PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

JUN 22 1936

VOL. XXVII APRIL, 1936 NO. 4

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

The Scoring of the Rearrangement Test 241
HERBERT S. CONRAD

The Learning of a Long Series 253
ISRAEL NEWMAN

In Criticism of the Detroit Tests of Learning Aptitude 258
D. A. WORCESTER AND STEPHEN M. COREY

A Physiological Theory of Thinking 263
W. H. PYLE

Judgments of Voice and Personality 272
ROSS STAGNER

Interest Correlates of Superior and Inferior Achievement in General Psychology 278
GRAHAM B. DIMMICK

Parental Occupation and Mental Ability 284
RUTH BYRNS AND V. A. C. HENMON

Evaluative Attitudes and Academic Success 292
J. W. M. ROTHNEY

On the Loss of Knowledge by Junior High-school Pupils over the Summer Vacation 299
MARIE SCHREPEL AND H. R. LASLETT

On the So-called Correlation Coefficient 304
MAURICE FRÉCHET

The Subjective Element in the Measurement of Relationship 307
PAUL HANLY FURFEY AND JOSEPH F. DALY

Book Reviews 310

$6.00 per Year · Published Monthly September to May

WARWICK & YORK, INC.

BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Established 1910

EDITORS

JACK W. DUNLAP
Fordham University

HAROLD E. JONES
University of California

PERCIVAL M. SYMONDS
Teachers College,
Columbia University

H. E. BUCHHOLZ, *Managing Editor*

THE price of the Journal is $6.00 a year in the United States; $6.40 in foreign countries. Part-year subscriptions are 90 cents for each number ordered. Back volumes are $7.00 each; back issues are $1.10 each except when more than five years old, and then $1.20 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Unsolicited manuscripts should be accompanied with return postage. Manuscripts, books and other materials for review, and correspondence regarding editorial and business matters should be addressed to the Publishers.

WARWICK AND YORK · *Publishers* · BALTIMORE, MD

THE ACTIVITY MOVEMENT

By CLYDE HISSONG

In an attempt to overcome the weaknesses of the traditional school organization many progressive schools have developed new programs. These programs are so similar in character that collectively the changes have been referred to as the activity movement. This movement has claimed the center of the educational stage for a length of time sufficient to have engendered widespread interest in its outcomes and in its basic philosophy.

In Doctor Hissong's study an attempt has been made to discover the principles underlying the present activity movement, to determine the influence of traditional concepts in shaping the trends of the movement, and to see if in the light of the present knowledge of the child and his relation to his environment the movement rests upon a justifiable basis.

$2.00 plus 10¢ postage

WARWICK AND YORK

Publishers BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXVII April, 1936 Number 4

THE SCORING OF THE REARRANGEMENT TEST

HERBERT S. CONRAD*

University of California

INTRODUCTION

In the scoring of objective-type questions, a correction for chance or guessing is usually recommended whenever the effect of chance may be considerable, and the evaluation of chance is reasonably accurate and easy. Such a correction is justified, because it significantly improves test-validity, with little (if any) loss of reliability.[7] An additional justification, on pedagogical grounds, is that class morale is weakened when the total score on a test is observed to reflect not only credits due to genuine knowledge, but also credits due to chance. For the rearrangement test (known also as the "continuity" or "sequence" test), no serviceable method of correcting for guessing seems yet to have been worked out.† The problem of the present paper is twofold: (1) To consider the effect of chance or guessing on scores in the rearrangement test; and (2) to suggest a table for the conversion of rearrangement test scores, into values that may be added directly to the scores from *other* types of objective items in a "mixed" test.

THE EFFECT OF CHANCE OR GUESSING

To receive any credit, a student's reponse to a test-question should be better than the response most probable by chance. The application of this principle to the rearrangement test is illustrated by the following

* For criticism of the manuscript and valuable suggestions, the writer is indebted to Dr. Harold D. Carter and Prof. Noel Keys.

† The method of scoring recommended by Cureton and Dunlap,[1] involving the computation of the rank-order coefficient of correlation between the correct arrangement and the student's, makes allowance for chance; but this method, as Odell[6] has shown, is too time-consuming to permit general use.

question, consisting of three items to be numbered in proper sequence (the question is taken from Tiegs, page 60[9]):

1 Battle of Bunker Hill
2 Declaration of Independence
3 Articles of Confederation

The events in this question occurred in the order in which they are listed above; the correct number of the events is, therefore, 1, 2, 3. Now, a student without the knowledge required to answer the question might number the events in any one of six ways (see Table I). Russell[8] and Odell[5] suggest that the student's score should be based directly on the difference between the *correct* numbering of the items of the rearrangement question, and the *student's* numbering.* The poorer the student's answer, the larger this difference becomes. Suppose, for example, that instead of numbering the items of the illustrative question 1, 2, 3, a student had numbered them 2, 3, 1 (as in column (5) of Table I); then the difference between the correct numbering and the student's would equal (as shown at the bottom of column (5′) of Table I) a total *discredit* of four points.

It is interesting to note, in Table I, that there are two *other* numberings, besides the suggested "2, 3, 1" which will yield a discredit-score of 4. There are two numberings which yield a discredit-score of 2; only

* This method of scoring differs from the methods recommended by Nesmith,[4] by Wilson[10] and Lang,[3] by John,[2] and by Cureton and Dunlap.[1] Odell[6] has shown that none of these alternative methods is sufficiently rapid to be practical. For this reason, the technique suggested by Russell and Odell will form the basis of the present article. The still simpler technique of merely crediting every correctly ranked item (the common "number right" formula) is rejected primarily for its failure to distinguish between small and large errors—a factor judged to be of some importance when the number of items in the rearrangement question is, say, four or more (3, 5, 8, 10).

The method suggested by Cureton and Dunlap[1] seems to have been accepted as the criterion by which to measure the validity of other methods of scoring the rearrangement question (6). We, however, consider the Russell-Odell method superior to that of Cureton and Dunlap, because the latter technique requires the *squaring* of the differences between the correct ranking and the student's ranking of the test-items. As Worcester[11] has pointed out, large discrepancies between the correct and the student's ranking are probably due to outright guessing. An outright guess may be fortunate or unfortunate; but in any event, if unfortunate, the chance discrepancy (from the correct ranking) does not seem to deserve so heavy a weighting as the process of squaring imposes. (Such a heavy weighting for an unfortunate guess of course implies—by the omission or reduction of penalty —a correspondingly heavy weighting for a more fortunate guess.)

one numbering, however—the correct one—yields a discredit-score of 0. (Discredit-scores of 3 and 1 do not occur.) These results are summarized in Table II. From Table II, it will be noticed that the chances are even (*viz.*, three out of six) that merely by outright guessing a student will earn a discredit-score of 4. From this consideration alone it is

TABLE I.—POSSIBLE NUMBERING OF THE THREE ITEMS OF THE ILLUSTRATIVE REARRANGEMENT QUESTION, AND POSSIBLE DISCREDIT-SCORES

Correct numbering	Possible numberings						Discredits					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(2′)	(3′)	(4′)	(5′)	(6′)	(7′)
1	1	1	2	2	3	3	0	0	1	1	2	2
2	2	3	1	3	1	2	0	1	1	1	1	0
3	3	2	3	1	2	1	0	1	0	2	1	2
	Discredit-scores.....						0	2	2	4	4	4

clear that, with a rearrangement question consisting of three items, a student whose discredit-score is as high as 4 should receive *no credit.*

TABLE II.—DISTRIBUTION OF DISCREDIT-SCORES BY CHANCE, FOR A THREE-ITEM REARRANGEMENT QUESTION

DISCREDIT-SCORE	FREQUENCY BY CHANCE
4	3
2	2
0	1
Total.....	6

By the construction of tables similar to Table I for rearrangement questions consisting of two, three, four, five, and six items, respectively, we have determined the frequency distributions of discredit-scores that are gotten merely by chance. These distributions are presented in Table III. In interpreting this table (which is similar to Table II), it is well to observe that any individual student's discredit-score will be an even number. (The method of obtaining the discredit-scores results in no odd or fractional values.) Looking at Table III, we see that when the number of items ("n") in the rearrangement question equals, for example, 4, the chances are thirteen out

of twenty-four, or greater than 50-50, that a student who guesses will make a discredit-score of 6 or worse. Hence, with a rearrangement question of four items, no student should receive any credit unless his discredit-score is *below* 6. Those whose discredit-score is 6 should receive zero credit; those whose discredit-score is 8 should also receive

TABLE III.—DISTRIBUTION OF DISCREDIT-SCORES BY CHANCE, FOR REARRANGEMENT ITEMS CONTAINING TWO, THREE, FOUR, FIVE, AND SIX ITEMS, RESPECTIVELY

Discredit-score*	Frequency by chance				
	n† = 2	n = 3	n = 4	n = 5	n = 6
18	..	..	..	..	36
16	..	..	..	..	100
14	..	..	..	..	136
12	...	..	..	20	148
10	..	..	..	24	137
8	..	..	4	35	93
6	..	..	9	24	46
4	..	3	7	12	18
2	1	2	3	4	5
0	1	1	1	1	1
Total.....	2	6	24	120	720
Mean.....	1.00	2.67	5.00	8.00	11.67
Median‡.....	1.00	3.00	5.22	8.09	11.81

* All discredit-scores are even integers; the method of obtaining the discredit-scores results in no odd or fractional values.

† "n" refers to the number of items in the rearrangement question.

‡ The median of a discrete series is not satisfactorily calculable. In computing these medians, we have assumed that a discredit-score of (say) 4 represents a continuum from 3.0 to 5.0, that a discredit-score of 6 represents a continuum from 5.0 to 7.0, etc.

zero credit (this is in accord with the method of scoring true-false items, where negative total scores are raised to zero*). Application of this

* It is possible to object to the raising of negative scores to zero, on the ground that a bad guess should receive less credit than a more fortunate guess. Such a suggestion fails, in the first place, to consider the question of class morale. Students should be taught to recognize when their knowledge is inadequate to support any answer; or if an answer is required (as it is, in the rearrangement test, since each item in such a test must be numbered), students should be made to believe that there is no profit to be gained in score, when there is no corresponding

method of scoring to the case where the number of items in the rearrangement question (n) equals 2, 3, 4, 5, and 6, respectively, results in the rule that zero credit should be assigned to all discredit-scores that are as bad as, or worse than, the following, respectively: 2, 4, 6, 10, 12. In this way, the scoring procedure safeguards against chance or guessing.

Our task now is to extend the method indicated above, to rearrangement questions containing more than six items. Suppose, for example, that there are ten items in the rearrangement question. There are over three million ways in which these ten items may be numbered; to be exact, the possible numberings are no less than $10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$, or 10!, or 3,628,800. Each of these numberings except one (the perfect answer) will differ to some extent from the correct numbering. To obtain the distribution of differences directly for the entire 3,628,800 numberings is obviously not feasible. The *mean* of the distribution of differences may, however, be exactly obtained by a simple observation. This observation is perhaps most obvious if we take a three-item rearrangement question as an example. Suppose the correct numbering of the items is 1, 2, 3. As already shown in Table I, there are six ways in which such a question can be numbered. In all these possible numberings (columns (2)–(7) of Table I), it will be noticed that the numbers 1, 2, and 3 each occur 3!/3 or $\frac{6}{3}$ or two times in each row. Consider the first item of the rearrangement question: This should be numbered 1. The student's numbering, however, if made by guess, will be 1, 2, or 3—and the differences between the correct and the student's possible numberings are, consequently, $|1\text{-}1|$, $|1\text{-}2|$, and $|1\text{-}3|$ (the differences are taken without regard to sign). Now, since each of the possible numbers, 1, 2, 3 occurs 3!/3 or two times in the first row, the *sum* of the *differences* for this first row equals $2|1\text{-}1| + 2|1\text{-}2| + 2|1\text{-}3|$, or 6. Similarly, the sum of the differences for the second item or second row equals $2|2\text{-}1| + 2|2\text{-}2| + 2|2\text{-}3|$ or 4; and the sum of the differences for the third item or third row equals $2|3\text{-}1| + 2|3\text{-}2| + 2|3\text{-}3|$ or 6. The *mean* of all the individual horizontal differences in all the rows is

$$\frac{2[|1\text{-}1| + |1\text{-}2| + |1\text{-}3| + |2\text{-}1| + |2\text{-}2| + |2\text{-}3| + |3\text{-}1| + |3\text{-}2| + |3\text{-}3|]}{18}$$

or $16\frac{6}{18}$ or .89.—Now, the "discredit-score" for any particular numbering of the three items of the rearrangement question is the *sum* of the discredits in any

---

knowledge to earn the score. The use of negative scores would yield a profit to those whose scores were merely zero—an undesirable outcome. Secondly, the reaction of students unfortunate enough to receive negative scores is likely to be too disgruntled and resentful to permit the development of the desired positive, constructive attitude toward test results. Third, it may well be doubted if negative scores would be understood by pupils in the elementary grades. On the whole, the suggestion to employ negative as well as zero scores does not seem worthy of consideration unless and until supported by evidence of improvement of test-validity or test-reliability.

column. The sum of the differences in the *average* column is merely three times the average horizontal difference (since each column consists of three horizontal differences, and the average column contains three average horizontal differences). For a three-item question this equals 3(.89) or 2.67—which checks exactly with the figure given as the mean chance discredit-score for a three-item question in Table III.

In general, the mean chance discredit-score for a rearrangement question consisting of n items is

$$\frac{n!/n[t_1|d_1| + t_2|d_2| + t_3|d_3| + \cdots]}{n!} = \frac{\Sigma t|d|}{n}$$

where

d is a difference (without regard to sign) between the correct numbering and the possible.

$\frac{n!}{n}$ is the number of times d appears in each horizontal set of differences (cf. text above).

t is the number of $\frac{n!}{n}$ sets in which each particular, given difference appears.

$n!$ (factorial n) is the number of ways in which a student could, by chance, number the items of the rearrangement question. Corresponding to each possible answer by the student, is a column of differences between the correct and the actual numbering of the items. The value of $n!$ is therefore the same as the number of columns of differences between the correct and the possible numberings.

The values of t and d may conveniently be obtained from a checker-board diagram, as indicated in Table IV (this illustration refers to a six-item question). In Table IV, the difference 1 occurs within the checker-board ten times, the difference 2 occurs eight times, etc. Application of the suggested formula gives

$$M = \frac{\Sigma t|d|}{n} = \frac{10(1) + 8(2) + 6(3) + 4(4) + 2(5)}{6} = 11.67,$$

which agrees exactly with the value given as the mean chance discredit-score for a six-item rearrangement question in Table III.

TABLE IV.—ILLUSTRATING A METHOD OF OBTAINING THE VALUES OF d AND t

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Applying the general formula, $M = \Sigma t|d|/n$, we have calculated the mean discredit-score that would be obtained by chance or guessing,

for rearrangement questions containing from two to twenty items.* These means are presented in Table V. In column (3) of Table V are listed the discredit-scores which—since they represent a performance as bad as, or worse than, the discredit-score to be expected by chance—*should be given a credit of zero.* For example, in a rearrangement question of four items, all discredit-scores of 6-8 receive a credit of zero; in a rearrangement question of eight items, all discredit-scores of 22-32 receive a credit of zero; etc. The designation (in column (3)) of the discredit-scores which should receive zero credit has been made with due regard to the fact that, as previously stated, odd-numbered and fractional discredit-scores do not occur for any individual case.

TABLE V.—MEAN DISCREDIT-SCORES BY CHANCE FOR REARRANGEMENT QUESTIONS CONSISTING OF FROM TWO TO TWENTY ITEMS; TOGETHER WITH DISCREDIT-SCORES WHICH SHOULD RECEIVE A CREDIT OF ZERO

n	Mean	Discredit-scores receiving a credit of zero	*n*	Mean	Discredit-scores receiving a credit of zero
(1)	(2)	(3)	(1)	(2)	(3)
2	1	2	12	47.67	48–72
3	2.67	4	13	56	56–84
4	5	6–8	14	65	66–98
5	8	8–12	15	74.67	76–112
6	11.67	12–18	16	85	86–128
7	16	16–24	17	96	96–144
8	21	22–32	18	107.67	108–162
9	26.67	28–40	19	120	120–180
10	33	34–50	20	133	134–200
11	40	40–60			

THE CONVERSION OF DISCREDIT-SCORES TO CREDITS

So far, we have determined only the discredit-scores which should receive zero-credit. The next problem is to convert the superior (better-than-chance) discredit-scores into positive credit scores. These credit scores should be such that they may be added directly to the credits on other types of question in the objective examination

* Judging from the distributions and statistical results given in Table III, we may infer that the fifty-percentile or median chance discredit-score would, for values of *n* above three, be close to the mean.

(such as true-false, completion, etc.). For this purpose, it is necessary first to decide on the general value of a perfectly answered rearrangement question. A convenient though somewhat arbitrary rule is that *a perfectly answered rearrangement question consisting of n items is* worth $n/2$ credits; thus, a rearrangement question of six items is worth 3 credits; of seven items is worth 3.5—we shall call this 4—credits; of eight items is also worth 4 credits; etc.* The next step is to apportion the positive credits to given discredit-scores in a well-graduated manner. In the case of a four-item rearrangement question, for example, our suggestion would be that a discredit-score of—

0 should receive a credit of 3
2 should receive a credit of 2
4 should receive a credit of 1
6 or 8 should receive a credit of 0

In the case of a seven-item rearrangement question, we would suggest that a discredit of—

0 should receive a credit of 4
2–4 should receive a credit of 3
6–8 should receive a credit of 2
10–14 should receive a credit of 1
16–24 should receive a credit of 0

It will be noticed that the gradation of credits for given discredits is not uniform: Thus, for the seven-item question, *three* discredit-scores (*viz.*, 10, 12, and 14) receive a credit of 1, but only *two* discredit-scores (*viz.*, 6 and 8) receive a credit of 2, and only *one* discredit score (*viz.*,0) receives the full credit of 4. The theory here is that as one approaches the zero-credit limit, the possibility that the student was guessing increases very rapidly (*cf.* the distributions in Table III); on the other hand, as one approaches the zero-discredit limit, the chances of guessing

* The proposal that a perfectly answered rearrangement question of n items should receive $n/2$ credits may be considered too conservative. One critic, for example, has observed to the writer that the difference between a correctly answered and an incorrectly answered true-false item (scored by the usual R-W (right minus wrong) formula) is two points. If rearrangement questions occurred only in connection with true-false items scored by the R-W formula, the proposed rule of $n/2$ credits would indeed be ungenerous. But rearrangement questions may more commonly be expected to appear in a "mixed" test with multiple-choice, matching, and completion items, as well as true-false. When the multiple-choice items are (as usual) scored one each, it would appear that the better rule for the rearrangement question is $n/2$, and, for the true-false items, $\frac{R\text{-}W}{2}$.

TABLE VI.—CREDIT SCORE CORRESPONDING TO EACH DISCREDIT-SCORE, FOR REARRANGEMENT QUESTIONS IN WHICH THE NUMBER OF ITEMS (n) VARIES FROM TWO TO TWENTY

Discredit-score	Number of items (n)																		
	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2
0	10	10	9	9	8	8	7	7	6	6	5	5	4	4	3	3	2	2	1
2	9	9	8	8	7	7	6	6	5	5	4	4	3	3	2	2	1	1	0
4	9	9	8	8	7	7	6	6	5	5	4	4	3	3	2	2	1	1	
6	9	9	8	8	7	7	6	6	5	5	4	4	3	3	2	2	1	0	
8	9	9	8	7	7	6	5	5	4	4	3	3	2	2	1	1	0		
10	9	8	7	7	6	6	5	5	4	4	3	3	2	2	1	1	0		
12	8	8	7	7	6	6	5	5	4	4	3	3	2	2	1	0	0		
14	8	8	7	7	6	6	5	4	4	3	3	2	2	1	1	0	0		
16	8	8	7	6	6	5	4	4	3	3	2	2	1	1	0	0	0		
18	8	8	7	6	6	5	4	4	3	3	2	2	1	1	0	0	0		
20	8	7	6	6	5	5	4	4	3	3	2	2	1	1	0	0	0		
22	8	7	6	6	5	5	4	4	3	3	2	2	1	1	0	0	0		
24	7	7	6	6	5	5	4	4	3	3	2	2	1	1	0	0	0		
26	7	7	6	5	5	4	4	3	3	3	2	2	1	1	0	0	0		
28	7	7	6	5	5	4	4	3	3	3	2	2	1	1	0	0	0		
30	7	6	6	5	4	4	4	3	3	2	2	2	1	1	0	0	0		
32	7	6	5	5	4	4	4	3	3	2	2	2	1	1	0	0	0		
34	7	6	5	5	4	4	4	3	3	2	2	2	1	1	0	0	0		
36	6	6	5	5	4	4	3	3	2	2	2	2	1	1	0	0	0		
38	6	6	5	4	4	4	3	3	2	2	2	2	1	1	0	0	0		
40	6	6	5	4	4	4	3	3	2	2	2	2	1	1	0	0	0		
42	6	5	5	4	4	3	3	3	2	2	2	2	1	1	0	0	0		
44	6	5	4	4	3	3	3	3	2	2	2	2	1	1	0	0	0		
46	6	5	4	4	3	3	3	3	2	2	2	2	1	1	0	0	0		
48	6	5	4	4	3	3	3	2	2	2	2	2	1	1	0	0	0		
50	5	5	4	3	3	3	2	2	2	2	2	2	1	1	0	0	0		
52	5	5	4	3	3	3	2	2	2	2	2	2	1	1	0	0	0		
54	5	5	4	3	3	3	2	2	2	2	2	2	1	1	0	0	0		
56	5	4	4	3	3	2	2	2	2	2	2	2	1	1	0	0	0		
58	5	4	3	3	2	2	2	2	2	2	2	2	1	1	0	0	0		
60	5	4	3	3	2	2	2	2	2	2	2	2	1	1	0	0	0		
62	5	4	3	3	2	2	2	2	2	2	2	2	1	1	0	0	0		
64	4	4	3	2	2	2	1	2	2	2	2	2	1	1	0	0	0		
66	4	4	3	2	2	2	1	1	2	2	2	2	1	1	0	0	0		
68	4	4	3	2	2	2	1	1	1	2	2	2	1	1	0	0	0		
70	4	3	3	2	2	1	1	1	1	1	2	2	1	1	0	0	0		
72	4	3	3	2	2	1	1	1	1	1	2	2	1	1	0	0	0		
74	4	3	2	2	2	1	1	1	1	1	2	2	1	1	0	0	0		
76	4	3	2	2	2	1	1	1	1	1	2	2	1	1	0	0	0		
78	4	3	2	2	2	1	1	1	1	1	2	2	1	1	0	0	0		
80	3	3	2	2	1	1	1	1	1	1	2	2	1	1	0	0	0		
82	3	3	2	2	1	1	1	1	1	1	2	2	1	1	0	0	0		
84	3	3	2	2	1	1	1	1	1	1	2	2	1	1	0	0	0		
86	3	2	2	2	1	1	1	1	1	1	2	2	1	1	0	0	0		
88	3	2	2	2	1	1	1	1	1	1	2	2	1	1	0	0	0		
90	3	2	2	1	1	1	1	1	1	1	2	2	1	1	0	0	0		
92	3	2	2	1	1	1	1	1	1	1	2	2	1	1	0	0	0		
94	3	2	2	1	1	1	1	1	1	1	2	2	1	1	0	0	0		
96	2	2	2	1	1	1	1	1	1	1	2	2	1	1	0	0	0		
98	2	2	2	1	1	1	1	1	1	1	2	2	1	1	0	0	0		
100	2	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0		
76–112	..	..	..	..	..	..	0												
86–128	..	..	..	..	..	0													
90–106	..	..	1	..	..														
96–112	2	..	..	..															
96–144	..	..	..	0															
102–118	..	1	..																
108–162	..	..	0																
114–132	1	..																	
120–180	..	0																	
134–200	0																		

(when $n = 6$ or greater) are practically *nil.* This deceleration in the effect of chance seems worth consideration, although we have not attempted to take exact account of it.

Table VI is a systematic presentation of the suggested credits to be assigned to stated discredits, for rearrangement questions containing from two to twenty items. Table VI incorporates the results in Table V; it is the only table that need be consulted in the process of scoring. Since most rearrangement questions consist of from four to ten items only, it is recommended that any one wishing to make use of Table VI should prepare a large, clear copy of the portion of the table from $n = 4$ through $n = 10$ (or such other portion of the table as fits his particular questions). Such a copy, placed before one when scoring rearrangement-test questions, will be found very convenient.

Practical convenience in the scoring of the rearrangement test can also be served by a little forethought when typing or mimeographing the questions. The column of dotted lines on which the student indicates his answer should (for right-handed scorers) be placed *about a half-inch to the left of the column of items.* This arrangement provides a convenient space for the scorer to record the difference between the correct answer (on the stencil) and the student's answer. An illustration of this arrangement is given below (the question is taken from Odell,[5] p. 409):

Number the following in order of *percentage of carbohydrate content, beginning with the largest:*

Correct Answer (Stencil)	Student's Answer	Difference	
	.....		sugar
	.....		potatoes
	.....		oatmeal
	.....		lima beans
	.....		rice
	.....		canned salmon

If the items of the rearrangement question are very brief, they may (for conservation of space) be typed in a row instead of a column, and the differences recorded in a space left below the student's answer. With this forethought in the provision of space for scoring, the recording and summing of differences is an easy task, and the use of Table VI becomes quite practical.

DISCUSSION

The fundamental assumption of the present paper is that no student's answer deserves credit, unless the answer surpasses that most probable by chance. The proposed scoring system assumes that the student will number the items of the rearrangement question *without omission,* and without such mechanical errors as using a given number

twice, or skipping a number. With regard to test-construction, it is assumed that no illegitimate or inappropriate cues are present—either in the directions or the rearrangement question itself—whereby students, without any knowledge of the subject-matter at hand, would on the average make an answer better than that to be expected by pure chance.

Although the present paper has been written with reference to rearrangement questions consisting of from two to twenty items, it is doubtful if the use of more than about twelve items in any one question is advisable. When the number of items exceeds twelve, the student is more likely to make merely mechanical errors in his numbering (such as using a number twice, skipping a number, etc.). Besides, the greater the number of items, the greater (in undue proportion) is the time required for answering, and also for scoring. Long rearrangement questions are also likely to be too easy, in the sense that the distinction between the items at the lower and upper extremes frequently becomes too obvious. Thus, a long rearrangement question covering an extensive span of civilization might start with "the discovery of fire," "the domestication of animals," etc., and run on up to "the age of steel," "the age of electricity," etc.; obviously, to distinguish the items belonging to the upper and lower extremes in such a rearrangement question is mere child's play. Finally, a very long rearrangement question is likely to provide a very intensive sampling of one topic or field, to the consequent neglect (since examination-time is always limited) of other equally important subject-matter.

SUMMARY AND CONCLUSIONS

The present paper considers two problems in the scoring of the rearrangement test: First, the proper allowance to be made for chance or guessing; second, the conversion of scores from rearrangement tests into values that may be added directly to the credits from other objective-type items. In answer to the first problem, Table 5 shows the discredit-scores which—since they represent a performance as bad as, or worse than, that to be expected on the average by chance—should be given a credit of zero. In answer to the second problem, a conversion-table (Table VI) is suggested. The conversions in this table take into account the decreasing probability that chance was operative as the discredit-scores approach zero. Since Table VI incorporates the results of Table V, it is the only table that need be consulted in the process of scoring.

Although Table VI has been prepared with reference to rearrangement questions consisting of from two to twenty items each, several considerations suggest that the use of a rearrangement exercise containing more than (about) twelve items is generally inadvisable.

The scoring method proposed in this paper represents an improvement over current scoring-technique for the rearrangement test. The suggested procedure is practical, particularly if, in the preparation of the rearrangement questions, space is definitely provided to record the numerical differences between the student's and the correct answer.

REFERENCES

1. Cureton, E. E. and Dunlap, J. W. "Scoring the Rearrangement or Continuity Test." *School Review*, 1930, Vol. XXXVIII, pp. 613–616.
2. John, Lenore. "A Comparison of Four Methods of Scoring the Continuity Test." *School Review*, 1930, Vol. XXXVIII, pp. 617–621.
3. Lang, A. R. *Modern Methods in Written Examinations.* Boston: Houghton Mifflin Co., 1930.
4. Nesmith, R. W. "Scoring the Continuity Test." *School Review*, 1929, Vol. XXXVII, pp. 764–766.
5. Odell, C. W. *Traditional Examinations and New-Type Tests.* New York: Century Co., 1928.
6. Odell, C. W. "Still More About Scoring Rearrangement or Continuity Tests." *School Review*, 1931, Vol. XXXIX, pp. 542–546.
7. Ruch, G. M. *The Objective or New-Type Examination.* Chicago: Scott, Foresman and Co., 1929.
8. Russell, C. *Classroom Tests.* Boston: Ginn and Co., 1926.
9. Tiegs, E. W. *Tests and Measurements for Teachers.* Boston: Houghton Mifflin Co., 1931.
10. Wilson, H. E. "Further Comments on the Scoring of Continuity Tests." *School Review*, 1930, Vol. XXXVIII, pp. 115–123.
11. Worcester, D. A. "Still Further Comments on the Scoring of the Continuity Test." *School Review*, 1930, Vol. XXXVIII, pp. 462–466.

THE LEARNING OF A LONG SERIES

ISRAEL NEWMAN

Augusta State Hospital, Augusta, Me.

According to Pavlov the execution of a habit consisting of a series of acts is based upon a conditioned chain-reflex mechanism. If so, each response (or the sum of the kinesthetic sensations plus the change in the situation resulting from the response) constitutes the stimulus for the next response.

There is, however, another possibility; namely, the continuous concurrence or recurrence of a process which acts repeatedly as a stimulus or the major part of one. Disturb the surface of water, as by the movement of a paddle, and waves are formed. The first wave gives rise to the second, the second to the third, etc. This corresponds roughly to the process of the chain-reflex. But there can be a series of waves as the result of wind continuously blowing over the surface of the water. This corresponds roughly to the proposition of a concurring or recurring inner process. We will discuss this second proposition more fully later.

It can be shown that the conception of a chain reflex is untenable. Whatever the number of the repetitions required for the retention of a series of facts, as a series of six nonsense syllables, that number of repetitions is insufficient for the retention of a longer series, say one of sixty syllables. If a given number of repetitions links, by modifying the synapses, *A* to *B*, *B* to *C*, *C* to *D*, etc., so that *A* becomes the stimulus for *B*, *B* for *C*, etc. up to six, why is the same linkage insufficient when the series is extended to sixty units? The question remains the same whether you consider the linkage to be synaptic, chemical, electrical, metaphysical or what-not.

G. E. Muller found that "A series of figures, consonants, syllables and so forth, is learned essentially by our combining the members of the series in an attitude of synthesis, so that they become a solid groups."[1] Granting that this is so, the question still remains unanswered. The most that we can infer is that it is harder to synthetize a longer series; but this does not explain why it is so. To say that the "energy" required for the synthesis of the short series is insufficient for the synthesis of the longer series were incorrect, since in repeating the longer series the same number of times there is the application of proportionately more energy. If in the repetition of six syllables six

seconds work is put in, in the repetition of sixty syllables ten times as much work is put in; yet in the latter instance the accomplishment is far from being ten times as great.

To answer this question we must go into the details of synthesis of sub-wholes into configurations. And it is not only this problem; there is also another fact inexplicable without the plausible assumption to be outlined. If the conditioned response depends upon the modification of special synapses, how does it happen that after one is conditioned to respond to letters one-fourth of an inch in diameter one's response is the same to letters of any other size involving different synapses?

It is obvious that the response is to the configuration and not to the sum of the sensations. But how does it happen that either of the great number of possible sensation-groups results in the synthesis of the identical engram which is the neuronic substratum of the configuration? If not so, how does it happen that any configuration of the same type, say that of a triangle or a group of letters, always leads to the same response?

Our assumption is that the sub-wholes of a configuration are non-sensory relation-experiences based upon the processes of special neurones which, for convenience sake, we will call *parasensory*. We can prove this by the following simple method:

When I sound two notes, *K* and *L*, in succession, I am conscious of three experiences: Of the lower note *K*, of the higher note *L* and of the (relation of) difference between them—of two sensory and one non-sensory experience. The "sense" of relation is more allied to a feeling in so far as it is too elusive for introspection. That it is experienced is certain, else one would not report such relations as difference, alikeness, between-ness, etc. If there are such experiences there are corresponding parasensory neurone processes.

Suppose we find that the most that can be appreciated are the differences between the sounds of *K* and *L*, *L* and *M*, *M* and *N*, *N* and *O* and so forth up the scale: The inference is that between the central sensory neurones corresponding to *K* and *L* there is a parasensory neurone interlinked with them so that it is stimulated when *K* and *L* are excited in crescendo succession. This parasensory neurone "registers" one unit of difference (*U*-1). There is obviously a similar *U*-1 neurone between *L* and *M*, etc. It is also evident that in addition to the above there are—between the sensory neurones—parasensory neurones registering greater degrees of differences, as *U*-2, *U*-3, *U*-4, etc. We will take no trouble to decide whether the latter are in relation with the sensory or the

parasensory (*U*-1) neurones. But for the sake of greater completion we will point out that, since one appreciates also difference in the diminuendo order, the inference is that all the above are duplicated.

It is out of such units of relation that the melody (the gestalt) is synthetized.

In the above system we note the fact that a given "quantum" of difference (or any other relation) may originate in more than one locality. Accordingly, the same melody, since its specific gestalt depends not upon the sensory elements but upon the interrelating units, may be the consequence of the stimulation of more than one set of sensory neurones.

Finding that the response is the same no matter what the sensory origin of the configuration, the conclusion is that there is a special center for each kind and degree of relation, as a center in which all the *U*-1 neurones terminate, another center in which all the *U*-2 terminate, etc.

A similar system can be diagramatically outlined for any other sensory field. (The writer has made a similar diagram also for time relations.)

If the above conclusions are true, the ultimate parasensory centers correspond to the higher regions of the brain. In instances of disturbances of these centers we would expect retention of at least the cruder elements of sensation with the loss of capacity for differentiation and for the appreciation of other relationships. These are precisely the results which Head[2] found in cases of cortical sensory disturbances.

Further evidence for the existence of special parasensory centers can be found in the following clinical data:

Riddoch[3] reported the case of a patient with occipital injury, who was unable to segregate objects from their surroundings. Holmes and Horrax[4] reported a case of cerebral injury which resulted in the inability to perceive interrelations of objects. This patient also had the difficulty at spatial orientation. The cases studied by Bender[5] showed impairment of the mechanisms concerned with the more complex integrations, while there was preservation of the capacity to integrate simpler and more elemental configurations. In fact, every case of sensory aphasia can be described as a case of parasensory paralysis.

On the basis of these assertions, there is a "double entry" of facts: One as a system of sensations, the other as a system of interrelations. The latter may be abstracted and manipulated by the mind

as a special entity, a *concept.* I thus may have the concept of a table without the sensory image of any particular table. What I grasp in such an instance is a system of interrelations.

With the above in mind we can explain why the memorizing and retention of the longer series requires more repetitions; in other words, why it is harder to synthetize a greater than a smaller number of units. Synthesis depends upon relation. If the relation is that of difference, the greater the amount of difference, the greater the readiness and the effectiveness of the synthesis. The round pattern amidst the many square patterns in the rug on the floor I isolate and synthetize most readily because of the greater degree of difference between it and the other patterns. Whatever it depends upon, the greater number of parasensory neurones stimulated or some other factors, it is always found that the greater degree of difference is the equivalent of greater intensity. Greater intensity means greater modification of the synapses. And it can be shown that, as compared with the longer series, the shorter series has higher degrees of difference. Suppose I have before me on the ground a row of ten objects, say, ten stones, differing but little in shape and resembling each other in size. The difference between the first and what is beyond it (the background) and the difference between the last and what is beyond it are greater than the difference between any other one and those on either side of it. Hence, the end-objects are more conspicuous, are better isolated. In the case of a row of only three objects, the central one is quite clearly distinctive as the middle object between the two conspicuous ones. But in the case of four objects, the difference between the two central ones is much smaller than the above, since the distinguishing feature of either is its being next to a conspicuous one. Thus we see that the greater the number of units in a row, the less the distinctions (the degrees of difference) between the units away from the ends. If any unit in a row differs from the others because of some special quality, such as size or color, it is correspondingly more conspicuous.[6]

The same seem to be the factors in a series to be memorized. The end units, because of their greater difference from what preceded and followed, respectively, are the easiest ones to be memorized. The intermediate ones are harder to memorize, depending upon their positions in the series.

We may say that in the execution of a habit consisting of a series of acts, there is either the continuous concurrence or the repeated recurrence of the concept (gestalt) of the procedure or intention as a whole.

This process, plus the one corresponding to the gestalt of the situation as it developes after every step, together constitute the (gestalt of the) stimulus. For instance: The streets in my town run East-West and North-South. I am in the habit of walking to my place of work every morning—to a factory located one half a mile Southeast of my home. To get there I zigzag my course: Along some streets I walk eastwards; along others—southwards. My concept (gestalt) of the general direction is either continuous or recurring at the street intersections. This concept plus the gestalt of the intersection together constitute the stimulus for the eastward or the southward turn, as the case may be. Introspection confirms the above assumption. And there is evidence for it in many experimental data too extensive to be outlined here.

The same holds true for the repetition of a memorized series of syllables: The (non-sensory) gestalt of the whole is part of the stimulus for each successive step.

Accept the above and one can explain the nature and the workings of instincts. The nucleus of an instinct is a concurrent or a recurrent non-sensory process closely associated with an emotion, a process which constitutes (the main) part of the stimulus, the rest being the other elements of the situation.

REFERENCES

1. Quoted from *Gestalt Psychology*, by W. Kohler, 1929, p. 285.
2. Head, H.: *Brain*, 1918, pt. II.
3. Riddoch, H.: "Dissociation in Visual Perception." *Brain*, Vol. XL, p. 14.
4. Holmes, G. and Horraz, G.: "Disturbances of Spatial Orientation, etc. *Arch. Neurol. & Psychiat.*, Vol. I, April, 1919, p. 385.
5. Bender, Lauretta: "Disturbances of Visuo-motor gestalt Function," etc. *Arch. Neurol. & Psychiat.*, Vol. XXX, 1933, p. 514.
6. See *Gestalt Psychology*, by W. Kohler, p. 305.

IN CRITICISM OF THE DETROIT TESTS OF LEARNING APTITUDE

D. A. WORCESTER AND STEPHEN M. COREY

University of Nebraska

When a new psychological test is published, educators and clinicians have a right to expect that the experience accumulated during the past thirty years should be used to guarantee a good test. Particularly is this true when the new test is presented with the evident expectation that it will supplant one already in use. Recently, there has come from the press the *Detroit Tests of Learning Aptitude*,[1] accompanied by the slogan "America to the Fore!,"[2] apparently a chauvinistic slur upon the European origins of the Binet-Simon Scale, and described with becoming modesty as a "long step forward in intelligence testing!"[2,3] An examination of certain materials appearing in the *Examiners' Handbook* has led the writers to question seriously the terms "long" and "forward."

The introductory discussions in the *Handbook* seem to have been written hastily, and with an organization that is sometimes baffling. On page 2, for example, the authors state that the new test requires a somewhat longer time for administration than does the Stanford Binet Tests, and then they add this comment: "This increase in testing time for younger children may disturb the somewhat prevalent idea that intelligence of young children improves under favorable environmental influences. Such may be an idle will-of-the-wisp, arising merely from inadequate testing." In just what way *the difference of a half-hour in testing time* is related to the problem of *increasing intelligence under favorable environmental influences* is not made particularly clear, nor in all probability will those who have investigated the influence of environment wish to be dismissed quite so cavalierly.

Part II, called "Diagnostic Interpretations," rightfully begins with the admonition that the *Detroit Tests of Learning Aptitude* do not measure everything, despite the nineteen separate scales. "Habits of industry, health, home environment, personality traits and many

[1] Baker, Harry J. and Leland, Bernice: *Detroit Tests of Learning Aptitude: Examiners' Handbook*, Bloomington, Ill.; Public School Publishing Company, 1935, pp. 157.

[2] The exclamation point *not* ours.

[3] It is recognized that the authors of the tests probably are not responsible for the publisher's "blurb," but they cannot be wholly free from its influence.

other factors all contribute their effects to the learning situation. The *Detroit Tests of Learning Aptitude* do not attempt to measure these other factors but chiefly that of the mental phases of learning." (P. 6.)[1] After this auspicious beginning, however, the authors present a very clear tabular description of "the entire series of tests with specific abilities which are related to success in the given tests" (p. 8)—*without any supporting evidence.* For example, it is claimed that the ability to "reason and comprehend" is measured by the following tests: (1) Pictorial absurdities, (2) verbal absurdities, (3) social adjustment, (4) orientation, and (5) broken pictures. "Practical judgment" on the other hand is measured in part by tests (1) motor speed and precision, (2) oral commissions, and (3) oral directions, *none* of which measures "reasoning or comprehension."

This nicety of distinction between "reasoning" and "judgment" is difficult to follow, particularly in the absence of any statistical evidence whatever—not even a coefficient of correlation. We have these matters on the authors' assertion only, and this is hardly sufficient for purposes of evaluation. Nor is there any reference made to articles in the learned journals, or in books where a conscientious psychologist might find out what the *Detroit Tests of Learning Aptitude* really do measure. This table, moreover, is entitled "The Tests and Specific Mental *Faculties,*" (italics ours) which of course implies, to some mental testers, an archaic psychological theory, the abandonment of which many believed to have been responsible for the success of Binet's method.

Part III of the manual is not in any sense a *profound* "Analysis of School Subjects." We are told in this section that reading, spelling, literature, English, and the foreign languages "make strong demands on verbal ability and adequate vocabulary" (p. 15), that motor ability is necessary for laboratory work, and verbal ability for reading textbooks. Perhaps such statements are intended for teachers whose intelligence the writers of the manual do not respect, although the warning that the simplicity of the Detroit technique is not an invitation for those without psychological training "to launch blindly into a testing program" (p. 3) hardly supports this hypothesis.

It would seem, too, that some evidence should have been presented when the claim is made (p. 16) that "in these subjects (health and physical education) motor skill is a first requisite." Especially may

[1] The construction of the second sentence is rather maladroit, but the meaning is clear.

one be allowed to question whether motor skill be a *first* requisite to health when the tests for motor skills pertain to such abilities as (1) making crosses in circles, (2) "to *show* me the window," "*open* the door, *then put* a mark on this paper, *then bring* me that book, *then stand* by the window," (3) to reproduce simple geometric designs which have been exposed for a short length of time, and (4) to "put a one in the circle and a cross in the square box," "draw a circle around the pig, a line under the apple, and make a cross on the cow." It will probably be somewhat difficult to convince anyone not a psychologist that these last tests measure either health or motor ability in any significant degree.

Part IV of the manual, "Statistical Evaluations," is most interesting in its lack of statistical validity. For example, all the items in the tests are arranged in order of increasing difficulty, "the final arrangement (having been based) upon the results of testing *approximately* 50 *average* children at each age for which the material is suitable." (P. 19.) Then, "if several hundreds or thousands had been tested there might have been a few minor changes in final order." (P. 19.) Very likely so! The only hint in this chapter of the criterion for "average" children is found on page 22. "Fifty children of average ability were selected at each age through a range at which the tests *appeared* suitable." (Italics ours.) "Criteria of average ability were the group mental ratings of $C-$, C, and $C+$ representing IQ's of ninety to one hundred ten, among children who are in grades for their age." Apparently the test was standardized by comparison with itself.

Regarding test intercorrelations, the authors write with devastating scorn: "Certain traditions and convictions have developed about the aristocracy of *high* correlations which are supposed to exist between various tests given to the same individuals. Any tests which did not fit into this elite family, having either *low* or actually negative correlations, have been frowned upon as black sheep." This tradition has either escaped the writers' notice or is of very recent origin. Intelligence testers have sought for low intercorrelations and high correlations with an outside criterion for almost twenty years.[1]

According to the authors, sixteen of the nineteen sub-tests have been used in intercorrelation computations. Although this information could have been presented very readily in a table, with great

[1] Yoakum, C. S. and Yerkes, R. M.: *Army Mental Tests.* New York: Henry Holt & Co., 1920, p. 6.

benefit to anyone trying to understand the nature of the scales, such was not done. Rather, three very general summary paragraphs were given, the data in which were based on test results from "approximately one hundred children ages eight to twelve years." (P. 20.) No probable errors were given nor any partial or multiple correlations.

Actual reliability data are published for but one sub-test out of nineteen—"the auditory attention span for unrelated words had a reliability of +.78 on one hundred fifty average children at an interval of three months." (P. 21.) For the tests as a whole, (p. 21) "the *Detroit Tests of Learning Aptitude* have had reliability determined for certain tests, but for the whole series of tests announcement will be made within a short time." ("America to the Fore!" Sell the test first; determine its reliability later.) It is also stated (pp. 21–22) that because certain other investigators found reliability coefficients ranging from +.78 to +.933 on tests "*somewhat similar*" (italics ours) "there is every reason to expect a reliability of +.80 or higher will be found." (A long step forward!)

Although many of the tests in the battery are scored by the point rather than the right-wrong method, no data are given showing the comparability of the scores computed by various testers for the same subjects. This is, in the opinion of some experts, an important consideration. If a score is to mean anything, the users of the test should know at least something about the chances in favor of two examiners obtaining the same score with the same individual.

Norms for the Detroit Tests were established with fifty children of average ability (IQ ninety to one hundred ten) for each age. "In some instances ten or more of the nineteen tests were given to the same children, in other instances only two or three tests." (P. 22.) This technique makes it impossible to learn (from the Manual) the exact number of subjects upon which any one test item was standardized. No data are given which would enable the student to determine the range of scores for any *CA*. No correlations were reported with the Binet Simon Scale although the authors, confirmed in such practices apparently, state that "the results of the two tests are very closely comparable." (P. 24.)

In all fairness it should be noted that the tests here criticized have the appearance of being good tests. They are well put up, and their administration appears to have been very carefully worked out. It is therefore all the more difficult for the present writers to cloak their surprise that a test battery such as the Detroit Series is so prematurely

published. Presumably it is an excellent measuring device, but so to conclude calls for more faith than judgment. Even a casual comparison of the standardization of the Detroit Scales with the pains taken by Terman and his associates causes one to wonder about the "capacity of benefit from experience." As was stated earlier, in this day and age one has reason to expect that a new scale, "a long step forward in intelligence testing!," might be standardized, and its reliability and validity determined, before final publication, but such hopes are probably näive.

A PHYSIOLOGICAL THEORY OF THINKING

W. H. PYLE

Wayne University

In the *Psychological Review* for July, 1924, I published a physiological theory of learning. In the present article I wish to make such modifications and extensions of the theory as are necessary to explain the various processes involved in thinking. The term *thinking* shall here be used to cover all types of memorial or ideational experience.

First let us get before us the psychological facts to be explained. Fairly well established are the following:

(1) The parts of simultaneous perceptual experience are so connected that they tend to persist in the same form of organization in ideational or secondary experience. (2) The parts of immediately successive perceptual experiences become so connected that the re-arousal of the first part in ideational experience tends to re-arouse the second part. (3) There is a tendency for similar experiences to become associated in memory. (4) Events that stand together in the relation of cause and effect become closely associated in ideational experience. (5) There is a tendency for perceptual experiences to grow and become richer as they reappear in ideational experience. (6) With the passage of time there comes about considerable interrelation of all one's ideational experience. (7) However, as time goes on there is a tendency for all experience that survives to become more and more skeletonized. There are, of course, many other aspects of thinking to be explained, such as meaning, individual differences, and genetic development; but the foregoing are the more important facts known concerning the interconnections of the various thinking processes. It is with interconnections or organization that this paper is primarily concerned.

Such are the psychological facts. What hypotheses concerning brain action will explain them? Let us first consider the matter of simultaneous experience. Here it is clear that the parts of simultaneous perceptual experience are connected and form a unitary consciousness. Upon the appearance of one part in secondary experience, there is a tendency for the other part or parts to appear also. It is therefore evident that the brain action underlying the simultaneous perceptual experience forms a brain-pattern with all the parts interconnected in such a way that the re-arousal of any element of this

brain-pattern *brings back all the other parts.* Simultaneous brain activities must of necessity be interconnected to account for the psychological facts. But how? In some such way as follows: Each excitation of the brain resulting from the complex of incoming stimuli is not a *separate excitation* but a part of the whole brain activity. To make this possible, each separate focus or center of brain activity must establish a gradient, become a point of less resistance, toward which other simultaneous excitations flow. The strength of flow toward any focus varies directly with the relative intensity of the activity at that point. All simultaneous excitations, whether from one or more sense organs, form a brain-pattern, each focus of activity passing over into each of the others, except as noted below in the discussion of inhibition. To illustrate: Suppose we have three foci of excitation X, Y, and Z. X flows into Y and Z; Y, into X and Z; Z, into X and Y. To illustrate by a concrete case, suppose I meet and am introduced to a friend. Let my seeing him be represented by brain excitation X; my hearing his name, by Y; and my hearing his voice, by Z. The brain-activity pattern set up by the perceptive processes persists, forming the basis of memory. The next day I hear the man's voice and think of how he looks and of his name. Or I may hear his name and think of his voice and appearance. Either of the three parts of the impression will arouse the other two. Indeed, the brain activity may be thought of as constituting a unitary pattern. Although we speak of points or foci, in most cases, doubtless, wide areas of brain activity are involved. But our hypothesis holds, whatever the extent of brain activity.

Simultaneous associations are explained by our hypothesis as above illustrated. To account for the facts of successive association we have to assume differential neural gradients, as already set forth, and a temporal existence of excitations. To illustrate: Suppose I have experience A, then before A dies away, B; and before B dies away, C. It is found to be a fact that A, B, and C become serially connected in memory, A arousing B; B arousing C. Experiment has found that these three experiences are also connected in memory in reverse order, C, B, A, but more weakly. The explanation is that the brain excitations corresponding to experiences A and B are partially simultaneous. The latter part of A is co-extensive with the stronger maximal part of B; the latter part of B with the stronger maximal part of C. The reverse associations are weaker, because in co-existence B is stronger than A; and C than B. Successive association, then, has precisely

the same explanation as does simultaneous association. We have but to remember that a brain process does not suddenly begin and end, but has a temporal existence, making possible the overlapping (in time) of successive brain excitations. Brain excitations persist long enough to account for the weak associative connection of experiences that are not temporally adjacent but separated by some seconds or even minutes. Psychological experiments have demonstrated such connections, and there is physiological evidence of the persistence of brain excitations over a period of seconds or even minutes.

Similar experiences become connected because brain-activity patterns are re-aroused by an excitation which in some respects, as far as the brain mechanism is concerned, is like, *i.e.*, identical with, some part of the original excitation underlying the experience in question. To illustrate: I know a man *X*. I meet another man *Y* whose voice, or hair, or nose, or some other characteristic is physically like that of *X*. The *X* brain-activity pattern is therefore re-aroused; and the similar *X* and *Y* become connected in my memory. Troland has a clear statement of the matter. "The probability that any system of afferent excitations will exactly reduplicate a previous one is infinitesimal. In practice, the degree of duplication which is required in order that the appropriate neurogram should be re-aroused, is comparatively inexact. There is a certain range of variation of the afferent processes, within which the given neurogram will be set off. . . . The neurogram is aroused by its afferent currents, as a key opens a lock which it fits; but the same lock can be opened by many separate keys." And again (p. 193) "Closely similar patterns tend to excite anotomically identical neurographic tissues."[1]

To illustrate how words of similar meaning acquire associative connections, let us consider a child's experience. He sees a tall object and hears someone say it is tall. At another time, seeing the same object he hears some one call it high; so tall and high are connected through an intermediate (identical) brain excitation, *i.e.*, the excitation underlying the perception of the tree. So it is with big and large and all other words of similar meaning. It is probable that as an individual grows older, association becomes more and more a verbal matter, and words denoting all sorts of logical relationships become connected through the laws of simultaneous and successive association. Some such relationships are those of genus-species, part-whole, opposites, subject-predicate, predicate-object, etc. But all such language and

[1] F. T. Troland: *The Principles of Psychophysiology*, Vol. III, p. 18.

logical associations doubtless have the same physiological explanation, as the types already discussed.

The cause-and-effect association falls, for explanation, into the same class with successive association. In nature, effects usually immediately follow causes. I strike a match and the gas is lighted. I turn a switch and the light goes on. It is therefore evident that natural phenomena, standing in the relation to each other of cause and effect, produce temporally overlapping brain-activity patterns, and the conscious processes involved have the same explanation as given above for successive association.

The spread of experience and the interconnections of all the parts of experience are due to the fact that a group of simultaneous stimuli do not fall upon an impartial brain. Simultaneous brain excitations are not only interconnected, but may arouse one or more *previously formed neurograms* in the manner above used to explain the association of similars. As a consequence, we have associative spreads and interconnections not "expected," to use Titchner's expression.[1] When a brain pattern is re-aroused it is possibly never in precisely the same setting. As a result our experiences become inconceivably interconnected, so that we can pass from one item of experience directly or indirectly to all others that still have surviving neurograms.

The preceding cases and illustrations explain most of the main facts in the organization of experience. Various details, however, need special discussion.

If in explaining simultaneous experience, we mean to include the corresponding brain excitation of all simultaneous stimuli, it is not true that all parts become connected and so survive in memory. It seems that only the parts of experience that are in the focus of attention have a brain-activity pattern surviving as the basis of memory. The other parts of the experience seem not ordinarily to survive. This can be explained by assuming that the brain functions as an accelerating and inhibiting organ. It is not a passive but an active machine, as has been pointed out by various writers, notably Piéron[2] and C. J. Herrick.[3] It reacts to a group of simultaneous incoming stimuli in a selective

---

[1] E. B. Titchner: *Textbook of Psychology*, 1910, p. 385. "The series is impressed upon a brain which is already the seat of a vast concourse of associative tendencies; and the incoming stimuli may thus excite to full or partial activity some pre-existing tendency whose arousal was neither expected nor intended."

[2] H. Piéron: *Thought and the Brain*, 1927, pp. 142, 234.

[3] C. J. Herrick: *The Thinking Machine*, 1926, p. 268.

fashion, accelerating some, forming thereby a pattern which in general alone is effective for memorial survival, while all other stimuli are more or less inhibited or shunted off from the main brain-activity pattern. The attended-to processes on the psychological side are probably the brain-activity pattern on the neurological side. Getting into the attention is a matter of getting into the brain pattern. The selective agency is the brain with its activity patterns of the moment. If these assumptions are true, attention, facilitation, and inhibition all have their explanation in brain-activity patterns. As time goes on, after an experience, the surviving brain-pattern becomes more and more skeletonized due to weaker parts of the pattern dying away, although, as explained above, it may have acquired new associates, centrally aroused.

How, it may very well be asked, does the brain facilitate and how does it inhibit? The experiments of Brown[1] and of Brown and Sherington perhaps have the closest bearing on this point. According to Brown not only do simultaneous excitations mutually flow into each other but they mutually facilitate each other. Sometimes, however, simultaneous processes mutually interfere with one another. To quote Brown (p. 141) "The double action of cortical points which is assumed above is one of 'mutual excitation' of activity, but it must be remembered that a double action of 'mutual inhibition' is equally possible between cortical points." If this mutual inhibition or interference of simultaneous processes is a fact, it probably has its explanation in the principle of chronaxie (Lapicque). Two processes interfere because they are not attuned and cannot act together. Child[2] in his discussion of this question says: "He (Lapicque) showed that an electric current of a certain intensity must act for a certain length of time to excite and on this basis he established a chronological coefficient which he called 'chronaxie.' This coefficient differs in different nerves and muscles and with different conditions and this difference, heterochronism, is of great physiological importance. And finally, he maintained that all these features of excitation can be interpreted in terms of electrical polarization of semipermeable membranes."

We have the problem of explaining how excitations of dissimilar chronaxies can become associated. Piéron has the following suggestions: "Thus even with prolongations intricated in many con-

[1] Graham Brown: *Quarterly Journal of Experimental Physiology*, Vol. IX, p. 117; Vol. X, p. 97.

[2] C. M. Child: *Physiological Foundations of Behavior*, 1924, p. 177.

nections, the excitation of a neurone will involve a selective response by predominant action on another neurone which is sufficiently isochronic: And so we get a reflex circuit. But if the exciting impulse reaches great intensity, responses will be obtained in more and more heterochronic neurones, which corresponds to the laws of the irradiation and generalization of reflexes."[1] It is thus probable that through more intense excitation or long continued or repeated excitations heterochronic neurones may be brought into simultaneous excitation.

The brain apparently is a vast system of interconnected nerve-elements which seem to be in a constant state of strain or tension (polarization?), tending toward equilibrium. Incoming sensory stimuli are continually upsetting the equilibrium. The tensions are constantly readjusted. In its readjustment, the brain acts more or less as a whole, or at least in patterns extending over large areas. *All the parts participating in the readjustment are mutually involved in an interchange of excitations.* The whole excitation pattern forms a neurogram with parts so interrelated that the re-arousal of any part sets off the whole complex process. The interrelationship within the neurogram is strengthened by repetition and weakened by disuse and the lapse of time. In the whole process of establishing a neurogram, the brain acts in a selective fashion, facilitating and inhibiting, as already explained. Inhibited excitations, as a rule, are not effective in memory because they have not become a part of a neurogram. The selective action of the brain at any moment probably depends upon tensions and organizations resulting from previous activity and also from heredity. At the beginning of an individual life such facilitating and inhibiting activities as the brain shows doubtless depend upon differential resistances in the inherited brain structure. With the growth of experience the excitation-transmission patterns formed in the brain become the bases of facilitation-inhibition activities. These would, of course, be different for every experiencing individual.

The hypotheses here set forth as a physiological explanation of the basis of organization of our experience has been in part suggested by various writers, but not wholly worked out by any one. M. F. Meyer's[2] theory, as I understand it, is that a maximal excitation drains or swamps the other simultaneous excitations, but this does not explain the mutual interchange of excitations, which is necessary to

[1] H. Piéron: *Op. cit.*, p. 139.

[2] M. F. Meyer: *The Fundamentals of Human Behavior*, 1911, pp. 233–237.

satisfy the facts. The Gestalt[1] idea of pattern and organization suggests the present theory, but no Gestaltist, so far as I know, has given any details of brain activity to explain their patterns. Indeed, Köhler repudiates the law of contiguity (simultaneity) and seems to think there is something in the *nature of ideas themselves* that underlies association. To the present writer such a notion is just the opposite of the truth. Any ideas whatsoever can become associated through simultaneous experience. Among the neurologists, I may cite Herrick[2] as having a clear conception of the situation. He says: "That which we know subjectively as the association of ideas may be pictured as involving neurologically the discharge of nervous energy in the cortex between two systems of neurons which have in some previous experience been physiologically united in some cortical reaction." He does not show, however, how they become "physiologically united." It is to explain this point that the theory set forth in this paper is proposed.

The crucial point in the hypothesis is that simultaneous centers of brain excitation mutually flow into one another and reënforce one another. The ultimate explanation of such a phenomenon would have to be found in the activity of the neurone. The nature of the activity of the neurone and the manner in which nerve processes are conducted are not known. What is probably the most widely accepted theory we may call the polarization-membrane-permeability theory. According to this theory excitation and transmission are matters of polarization and the movement of ions through the outside membrane of the nerve. Many investigators have had a part in the elaboration of this theory. Perhaps Lillie has given it the fullest expression. Brief statements with illustrations may be found in Troland.[3] But why should simultaneous excitations mutually flow into one another?

George T. Johnson[4] gives expression to a view, which, if true, would explain the assumptions of our hypothesis. He summarizes his view as follows:

> To briefly review, we would believe that the functional unit of the nervous system is the neurone and the fundamental function of the neurone is to react to stimulation by the production of excitation waves in the axis cylinder

[1] W. Köhler: *Gestalt Psychology*, 1929, pp. 278–282.

[2] C. J. Herrick: *Introduction to Neurology*, 1921, p. 329.

[3] L. T. Troland: *Op. cit.*, Vol. II, p. 24; Vol. III, p. 47.

[4] Geo. T. Johnson: "A Survey of the Physiology of Cerebration." *Jr. of Abnormal Psychology and Social Psychology*, Vol. XVI, pp. 115–136.

process: That the spread of activity is determined by the neurone threshold and *stimulus* response ratio of the different neurones, which depends on the irritability of the different synapses. Further that the irritability of cortical synapses is rapidly increased by repeated activity (producing a lower neurone threshold and *stimulus* response (ratio)) and finally that activity in neurones lowers the resistance of their afferent synapses and the resistance of the different efferent synapses influences the spread of the excitation wave over the different collaterals of neurones and a lowered resistance of the efferent synapses facilitates activity in short neurones.

Herrick[1] expresses a similar view as follows:

It is well known physiologically that a nervous circuit which is in action is prepotent over resting circuits, and when a response is in process intercurrent stimuli tend to affect the acting circuit rather than to activate resting systems of neurons which would otherwise be excited. This is usually ascribed to the greater permeability of the active systems, whose internal inertia has already been overcome so that they tend to capture or draw off available nervous energies from surrounding parts.

These two quotations show that the type of brain action demanded by the hypothesis of this paper is acceptable to the neurologist. The rat decerebrating experiments of Lashley[2] have raised some doubts as to the possibilities of brain localization, but a certain degree of functional cerebral localization is necessary to account for human memory and the various facts of mental organization. It may, however, be a very different type of localization from that which has heretofore been supposed. A cerebral pattern may involve widely different parts of the cortex and innumerable association fibres; and a certain amount of shifting and substitution may be functionally possible. Lashley himself has some idea as to the necessities and possibilities of the situation as shown by the following quotation (p. 173):

There is no evidence to support this belief in identity of nervous elements. On the contrary, it is very doubtful if the same neurons or synapses are involved even in two similar reactions to the same stimulus. Our data seem to prove that the structural elements are relatively unimportant for integration and that the common elements must be some sort of dynamic patterns, determined by the relations or ratios among the parts of the system and not by the specific neurons activated. If this be true, we cannot, on

[1] C. J. Herrick: *Brains of Rats and Men*, 1926, p. 57.

[2] K. S. Lashley: *Brain Mechanisms and Intelligence*, 1929.

the basis of our present knowledge of the nervous system, set any limit to the kinds or amount of transfer possible or to the sort of relations which may be directly recognized.

And it must not be forgotten that Lashley's experiments were concerned with the relative unorganized cerebrums of rats.

It is not our purpose to discuss here all aspects of thought processes. But the explanations of all psychological facts must be of a piece and hang together. The physiological explanation of individual differences in what is usually spoken of as capacity to think or reason may probably be looked for in the number and organization of association fibres, nature of synaptic organizations, amount of available cerebral energy, and the ability of the brain to function as an accelerating and inhibiting organ. For the facts we must await the future work of the psychologist and neurologist.

JUDGMENTS OF VOICE AND PERSONALITY

ROSS STAGNER

University of Akron

The pioneer work of Pear,[1] demonstrating the possibility that personality may be judged on the basis of voice alone with considerable accuracy, has been followed by studies employing more definite controls, *e.g.*, Allport and Cantril[2] and Taylor.[3] In these later studies, the task of the judge has been made more objective by supplying him with alternative choices as to personality descriptions, and there has been also an increasing specificity in the personality-judgments required. Pear's radio listeners wrote freehand descriptions of personalities. Allport and Cantril asked their subjects to check one of three (or more) alternatives, *e.g.*, very ascendant, slightly submissive, very submissive. Taylor required his judges to "describe" the possessor of each voice by checking items from a long list of adjectives, phrases, items descriptive of emotionality and introversion, etc.

In all of these studies, the voice itself has been treated as an unanalyzable whole. Allport and Cantril even go so far as to say that it would be "absurdly atomistic" to attempt to correlate specific types of vocal reaction to specific types of personality reaction. This position seems entirely untenable. In the first place, no attempt to correlate different aspects of human behavior is absurd. In the second, we are all aware of the fact that specific differences in vocal expressiveness are correlated very definitely with specific personality differences in extreme cases. Anyone who has heard the intense, fluent, emphatic voice of the manic patient, the low, dull, unintelligible monotone of the hebephrenic, or the "whine" of many neurasthenics, knows that personality is related to specific vocal characteristics.

The present experiment was intended to explore roughly the possibility that judgments of personality traits on the basis of voice might be shown to have a foundation in the judge's reaction to certain specific qualities of the voice. For this reason it was not considered necessary that we take elaborate precautions to exclude visual and other cues. If these cues operate in the judgment of the personality traits, then the correlation of judged *vocal* characteristics with these traits will approach zero. We are not concerned, as were the previous studies, with showing that personality *can* be judged from voice (in these, the exclusion of extraneous cues was essential); we are simply

seeking relationships among judgments, and if extraneous cues operate, they will only attenuate the observed correlations.

METHOD

Twenty-five students in the author's social psychology class served as subjects. Having finished the study of gesture and expressive movements as an aspect of personality, the lecturer pointed out to the class that voice might also have some relationship to personality traits, although no definite experimental evidence was availalbe. Graphic rating scales were then distributed and they were asked to judge, solely on the basis of voice, with no reliance on visual cues or previous knowledge of the subject, certain specific vocal characteristics and also the traits "aggressiveness" and "nervousness," terms with which they were quite familiar. Five students then read typewritten excerpts from a book on personality which they had previously practiced (to reduce reading difficulties). Five boys read at one class meeting, five girls at another. Three to four minutes reading was done by each "voice." These ten individuals also filled out two standardized questionnaires on personality, the Bernreuter inventory and the Wisconsin Scale of Personality Traits.

Seven graphic ratings were called for, as follows: (1) General impression; (2) voice intensity; (3) flow of speech; (4) poise; (5) clearness; (6) aggressiveness; and (7) nervousness. Each end of the scale was described by a phrase, and the favorable and unfavorable ends were alternated to reduce halo effects. Ratings were on an eleven-point scale (and some judgments were made on all eleven points of every scale).

Since we were interested in generalizations rather than in specific subjects, the intercorrelations of the rating scales were computed by sex groups rather than by individual subjects. In fact, the restricted range of some individual ratings would have made correlations by individual subjects difficult. The *ranking* of each subject, however, has been computed for certain correlations requiring individual scores.

RESULTS

There was no convenient way of determining the reliability of the ratings. To see if there were general agreement on the characteristics rated, the judges were divided into two equal groups and the ranking given each voice by group *A* was correlated with that by

group *B*. The results, as shown in Table I, indicate a considerable degree of consistency or "social agreement," as Taylor[3] calls it. Scales two, three, four and five (the so-called voice characteristics) give a median *r* of .90, while one, six and seven, the general characteristics, give a median of only .80. This seems to indicate that there will be greater social agreement on relatively objective characteristics like these, than on inferred characteristics such as personality traits. These "reliabilities" are exceptionally high because each single score is based on the average of twelve judgments.

TABLE I.—CORRELATION OF RANKINGS OF VOICES BY JUDGES IN GROUPS *A* AND *B*

	Five men	PE_r	Five women	PE_r
1. General impression	.70	.14	.90	.05
2. Voice intensity	.90	.05	.97	.03
3. Flow of speech	.80	.10	.97	.03
4. Poise	.70	.14	.97	.03
5. Clearness	.90	.05	.80	.10
6. Aggressiveness	.30	.25	.57	.18
7. Nervousness	.90	.05	.95	.03

The rankings of the five subjects in each group by all twenty-five judges on scales two, three, four and five were also correlated with the rankings on six and seven to estimate the degree of relationship. This method gives very high correlations, so high that we are inclined to be suspicious of them. They may be high because of the high reliabilities of rankings based on twenty-five judgments, but it is more likely that the use of the average rating, instead of the wide range of ratings given by all the judges, greatly increased the "halo effect." It is interesting, in view of Table II below, that correlations with nervousness are better than those with aggressiveness, only one exception appearing in eight coefficients.

Table II gives the correlations we are especially interested in. To obtain these, the reaction of each judge to each voice was treated as a single case, and the one hundred twenty-five judgments were correlated for males and females separately. Thus the relationships are not an average relationship for *voices*, but the relationships existing in the minds of the *judges* who, in rating the personality traits, may (or may not) have been influenced by specific vocal characteristics.

It is clear from Table II that the judges were not making judgments by chance, and that they were not simply alternating high and low checks indiscriminately. The halo effect is apparently at a minimum. These correlations seem to show definite relationships in the judges' minds as they reacted to the voices.

TABLE II.—CORRELATIONS BETWEEN JUDGES' RATINGS OF VOICE CHARACTERISTICS AND OF PERSONALITY TRAITS

	Voice intensity	Flow of speech	Poise	Clearness
Ratings of men on:				
Aggressiveness	.24	.28	.46	.29
PE_r	.07	.06	.05	.06
Nervousness	−.25	−.39	−.44	−.39
PE_r	.07	.06	.06	.06
Ratings of women on:				
Aggressiveness	.54	.35	.50	.38
PE_r	.05	.06	.05	.06
Nervousness	−.28	−.67	−.65	−.56
PE_r	.06	.04	.04	.05

As noted above, nervousness gives higher correlations than aggressiveness. This is probably due to the lower reliability of ratings of the latter (Table I) and this in turn to the fact that in *reading* a passage one is unable to manifest aggressiveness (although he would if arguing or debating).

Aggressiveness is related to voice intensity (.54, women) and to poise (.46, men, .50, women). Nervousness shows a respectable relationship to flow of speech, poise, and clearness. Poise is, strictly speaking, a generalized rather than a vocal reaction, and thus might be expected to correlate well with both personality traits. Both sexes show no relation between voice intensity and nervousness; likewise they show little correlation of aggressiveness with flow of speech and clearness. These consistencies seem of considerable importance.

As mentioned above, each of the "voices" took the Bernreuter and Wisconsin personality scales for "self-judgments" of personality. For the Bernreuter scale we used the scores F1-*C* and F2-*S* recently devised by Flanagan,[4] which he has shown give the real essence of the four Bernreuter scores. These two scores were used to rank the subjects and compute rank-difference correlations with the voice

rating. Of the sixteen correlations, only two exceeded .50 (with $N = 5$, we considered this a minimum of significance). These were both with the female subjects. Confidence correlated .80 with voice intensity, and sociability, −.50 with poise! Such inconsistency indicates that little can be expected from this approach.

The Wisconsin scale[5] has four scores, *W*, *X*, *Y* and *Z*, which may be roughly defined as follows: *W*, "emotionality" or nervousness; *X*, subjective orientation ("introversion"); *Y*, self-esteem; *Z*, social passivity or seclusiveness. *W* and *Y* have high positive correlations with voice ratings for men—and high negative correlations for women. *X* gives no significant correlations. *Z* gives high positive correlations for men and for "voice intensity" with women, the others being low. The correlation of self-judgments on both Bernreuter and Wisconsin scales with the pooled social judgments on aggressiveness and nervousness were low. From this miscellany of relationships we can only conclude with Taylor that there is no agreement between self-judgments and social judgments based on vocal cues.

DISCUSSION

The importance of social conditioning of various specific vocal responses has been emphasized by Sapir.[6] It is part of a particular cultural heritage, he points out, that one speak with a certain intonation, a certain range of intensity, a certain rate, and so on. Within the range of socially approved variation, individual differences enter, and these may be expected to show personality relationships. The success with which the individual fits himself into the socially approved scheme is also a personality variable. It should be plain, then, that not only the analyzed "voice as a whole" in all its gestaltian purity may be a cue for judgments of personality, but also these various specific aspects which are subject in varying degrees to social conditioning.

Our results are admittedly based on too small a number of voices and we have not determined accurately the reliability of the ratings of vocal characteristics. The use of a mechanical speaker, despite its distortion, might be a valuable control. Despite these criticisms, we believe that our results clearly demonstrate *the use of specific vocal cues in the process of making judgments about personality.* When this study has been amplified by persons skilled in speech analysis, it may lead to suggestions for improving certain kinds of voices which are not definitely defective, but which carry to listeners the inference of an unpleasant personality. Mental hygiene workers are familiar

with the value of public speaking training in relieving certain kinds of personal difficulties. This effect may be due as much to the change in the attitude of listeners (when this person speaks) as to the change in the subject's emotions. In fact, it is likely that this relationship exists only on the stimulus side; for, as our results and those of Taylor show, there is no consistent correlation between *self-judgments* of personality and social judgments of voice. The increased "self-confidence" of the speaker after such training appears to be less important than the improved reception he obtains.

SUMMARY

1. Ratings of voice and of personality traits by twenty-five judges of ten speakers are intercorrelated to show that the judge seems to be influenced by specific vocal characteristics in making his judgment of the personality traits.

2. This fact is intelligible in view of the social-conditioning theory of personality formation.

3. Mental hygienists may find (even in cases of normal speech) that a voice clinic may improve the social relationships of problem cases.

BIBLIOGRAPHY

1. Pear, T. H.: *Voice and Personality.* New York: John Wiley and Sons, 1931.
2. Allport, G. W., and Cantril, H.: "Judging personality from voice." *J. Soc. Psychol.*, Vol. V, 1934, pp. 37–55.
3. Taylor, H. C.: "Social agreement on personality traits as judged from speech." *J. Soc. Psychol.*, Vol. V, 1934, pp. 244–248.
4. Flanagan, J. C.: *Factor analysis in the study of personality.* Stanford University Press, 1935.
5. Stagner, R.: "The Wisconsin Scale of Personality Traits." *J. Abn. & Soc. Psychol.*, 1936 (to appear).
6. Sapir, E.: "Speech as a personality trait." *Amer. J. Sociol.*, Vol. XXXII, 1927, pp. 892–905.

INTEREST CORRELATES OF SUPERIOR AND INFERIOR ACHIEVEMENT IN GENERAL PSYCHOLOGY

GRAHAM B. DIMMICK

University of Kentucky

The purpose of this study was to determine the relationship between scholastic standing in general psychology, as represented by the final grade earned in the course, and the choices made in respect to the contrasting items on Miner's Blank for the Analysis of Work Interests.[2] Comparisons were made in regard to interests and work conditions between the psychology student group which earned a final grade of *A* or *B* and the group which earned a final grade of *D* or *E*. Statistically reliable differences obtaining between these two groups in respect to items on the blank would not only contribute to an understanding of the interests and work conditions which were the concomitants of satisfactory achievement in general psychology, but they would make such items available as criteria which might possibly be useful in sectioning general psychology students for the purpose of differentiating instruction. A differential checking of the items by a random group, in comparison with a representative sample of general psychology students with satisfactory achievement, might serve guidance purposes by thus revealing fundamental interest trends which should be of value in advising students in the selection of suitable fields of study.

The subjects employed in this study were two hundred fifty students who had taken the beginning course in psychology at the University of Kentucky. This course extends throughout the college year and there is required two hours attendance in recitation and laboratory respectively each week.

Each student, as a part of the regular routine involved in registering upon entrance to the University, checked each of the twenty-eight pairs of contrasts of interests and working conditions as they appeared upon the Personal History Record Blank. This division of the History Record Blank duplicates the items on Miner's Blank for the Analysis of Work Interests. Instructions to the student in regard to these items were as follows:

> Notice the contrasts in working conditions listed below. See how they are related to your own interests and abilities. Wherever you can distinguish a difference in yourself, check that one in the pair under which you believe

you would do the better work. Consider in which of the pair, after training, you would trust yourself more to meet the competition of others.

In respect to only one of the twenty-eight pairs of contrasts was the reliability coefficient less than .69. This coefficient was obtained by having the blank filled out by one hundred students three months after the initial performance, and it was calculated upon the basis of the percentage changing from one of the pair of contrasting interests to the other.

The possible range in grade in beginning psychology extends from *A* through *E*. The distribution of grades for the group of students

TABLE I.—PERCENTAGE OF SUBJECTS EARNING EACH GRADE

Grade	*A*	*B*	*C*	*D*	*E*	Total
Per cent	10.8	18.0	46.8	17.6	6.8	
N	27	45	117	44	17	250

employed in this study is presented as shown in Table I. The percentage of subjects checking each choice of interest or working condition for the twenty-eight items was then computed for each letter grade group.

Contrasting items	Per cent *A* and *B* checking	Per cent *D* and *E* checking	Critical ratio d/σ_d
( ) Thoughtful hand work			
( ) Thoughtful head work	83	66	2.24
( ) Indoor work	64	48	1.88
( ) Outdoor work			
( ) Slow movements			
( ) Rapid movements	51	67	1.91
( ) Organizing people	64	46	2.12
( ) Constructing things			
( ) Work requiring judgment	79	62	2.15
( ) Work requiring speed			
( ) Work requiring common sense			
( ) Work requiring specific information	51	18	4.07

Inter-group comparisons were made in terms of the percentage differences in checking the contrasts in interests and working conditions. These percentage differences were considered statistically reliable when the chances of their being real were ninety-seven in one hundred or above. The formula for obtaining the standard error of a proportion as given by Yule[4] was used in calculating the standard error of percentage differences.

Statistically reliable differences between the percentage of the combined *A* and *B* group and the combined *D* and *E* group were found for the items as shown in the table at the bottom of page 279.

Differences approaching statistical reliability were found for the following additional items:

Contrasting items	Per cent *A* and *B* checking	Per cent *D* and *E* checking	Critical ratio d/σ_d
( ) Working by yourself			
( ) Working with others	61	49	1.38
( ) Work requiring expression	57	69	1.46
( ) Work requiring computation			
( ) Work requiring artistic taste	61	46	1.72
( ) Work requiring artistic creation			

When the two extreme groups were compared, the *A* and the *E* group, reliable differences were found for the following items:

Contrasting items	Per cent *A* checking	Per cent *E* checking	Critical ratio d/σ_d
( ) Slow movements			
( ) Rapid movements	33	65	2.13
( ) Operating appliances			
( ) Repairing appliances	0	41	2.73
( ) Work requiring common sense			
( ) Work requiring specific information	56	24	2.79

Differences approaching statistical reliability were found for the following additional items in this comparison of the extreme groups:

Contrasting items	Per cent *A* checking	Per cent *E* checking	Critical ratio d/σ_d
( ) Methodical work			
( ) Meeting new conditions	89	65	1.71
( ) Working by yourself	30	53	1.53
( ) Working with others			
( ) Work requiring energy			
( ) Work requiring patience	52	71	1.27
( ) Thoughtful hand work			
( ) Thoughtful head work	89	71	1.41

A comparison of the psychology student group with an unselected or random college group of five hundred seventy students in respect to interests and preferred work conditions revealed statistically reliable differences for the items as shown in the table on page 282.

SUMMARY AND CONCLUSIONS

Seventy-two students who earned a final grade of *A* or *B* in the beginning course in psychology were compared with sixty-one students in the same course who earned a final grade of *D* or *E* in respect to the percentage of each group checking the contrasts in interests and working conditions as presented on Miner's Blank for the Analysis of Work Interests. The reliabilities for each of the percentage differences were determined. It was found that the *A* or *B* students differed from the *D* or *E* students in respect to their interests and the conditions under which they preferred to work, and that these differences were statistically reliable for six of the items on the blank. Differences which approached statistically reliability were found for three additional items.

Comparisons were also made in a similar manner between two hundred fifty psychology students and a random group of five hundred seventy college freshmen. Reliable differences were found for fourteen items.

Contrasting items	Per cent psychology group checking	Per cent random group checking	Critical ratio d/σ_d
( ) Thoughtful hand work			
( ) Thoughtful head work	74	62	3.30
( ) Skilled hand work	76	65	3.14
( ) Skilled heavy work			
( ) Indoor work	55	29	7.02
( ) Outdoor work			
( ) Organizing people	53	38	3.84
( ) Constructing things			
( ) Working by yourself	43	32	2.97
( ) Working with others			
( ) Work requiring judgment	83	66	5.48
( ) Work requiring industry			
( ) Large future success	84	75	3.00
( ) Smaller immediate success			
( ) Work requiring expression	71	47	6.67
( ) Work requiring computation			
( ) Increasing amount of production	15	22	2.33
( ) Improving quality of product			
( ) Working with things			
( ) Working with people	24	31	2.00
( ) Work requiring energy			
( ) Work requiring patience	45	54	2.31
( ) Making things as directed			
( ) Following own methods	30	22	2.42
( ) Work requiring artistic taste	49	39	2.56
( ) Work requiring artistic creation			
( ) Welfare work	42	35	1.89
( ) Taking part in entertainments			

Besides tested intelligence, there exist certain fundamental interests and attitudes which must be contributory to success in specific college courses. Such an assumption seems justified in attempting to account for the discrepancy between intelligence and achievement. This study has pointed out certain fundamental differences in interests and work conditions which seem to be the concomitants of successful achievement in beginning psychology. Moreover, differences in interest trends have been indicated for students completing the first course in psychology in contrast with a random group of college freshmen.

This study suggests the possibility of establishing criteria which may be utilized as a basis for guidance in sectioning students in the beginning psychology course. Certainly criteria which will be adequate for this purpose will necessarily include, along with an intelligence rating, some consideration of the individual's work interests and attitudes.

BIBLIOGRAPHY

1. Miner, J. B.: "An Aid to the Analysis of Vocational Interests." *Journal of Educational Research*, April, 1922, pp. 311–323.
2. Miner, J. B.: "A Method for Evaluating a Psychograph for Vocational Guidance." *Journal of Educational Psychology*, May, 1926, pp. 331–340.
3. Miner, J. B.: "An Analysis of Vocational Interests." *School Review*, Vol. XXXIII, 1925, pp. 744–754.
4. Yule, G. Udny: *An Introduction to the Theory of Statistics*. London: Charles Griffin & Co., Ltd., 1911.

PARENTAL OCCUPATION AND MENTAL ABILITY

RUTH BYRNS

Fordham University

AND

V. A. C. HENMON

University of Wisconsin

Numerous studies have been made of the relation of parental occupation to the intelligence of children, but none has been based on a sufficiently large number of cases to establish the hierarchy of parental occupations in relation to mental ability with requisite accuracy. The variability and overlapping between different occupational groups is so great that only a very large number of cases could determine whether there are statistically reliable differences between occupational groups.

Over one hundred thousand cases have been analyzed for this study and the reliability of the differences between groups has been determined. The data have been collected through the Wisconsin State testing program in which one hundred thousand eight hundred twenty high-school seniors have taken aptitude tests and listed the father's occupation on information blanks. The percentile scores on the aptitude tests were taken as the measure of scholastic or mental aptitude. The tests given the high-school students were the Ohio State University Psychological Test in 1929, the American Council Psychological Examination in 1930 and 1931, and the Henmon-Nelson Test of Mental Ability in 1932 and 1933.

Since the subjects of this study were all high-school seniors, differences due to unequal education and to language handicaps were cut to a minimum. The disturbing factor of selection, however, probably is increased rather than reduced by use of subjects of equal and advanced school rank. It is impossible to estimate the amount of selection within the various groups, but it seems likely that the selective process may not have been consistently rigorous in different groups. Interpretations and conclusions from these data therefore must be made with the greatest caution.

The parental occupations fall into nine general groups: Professional, executive, business, clerical, skilled, semi-skilled, unskilled, farmer, miscellaneous; the unemployed constitute a tenth group. This classification is an adaptation of the five-fold grouping suggested by

TABLE I.—PARENTAL OCCUPATION AND SCHOLASTIC APTITUDE AMONG ONE HUNDRED THOUSAND EIGHT HUNDRED TWENTY WISCONSIN HIGH-SCHOOL SENIORS

Psychological percentiles	Professional	Executive	Business	Clerical	Miscellaneous	Skilled
90–100	1,330	654	2,461	579	282	1,831
80– 89	885	593	2,150	462	247	1,831
70– 79	710	535	1,781	370	244	1,636
60– 69	603	485	1,859	372	264	1,797
50– 59	484	373	1,489	387	228	1,682
40– 49	562	488	1,754	452	273	2,080
30– 39	389	398	1,456	289	183	1,711
20– 29	378	293	1,246	281	201	1,554
10– 19	388	341	1,407	289	204	1,630
0– 9	299	253	1,205	258	179	1,418
Total	6,028	4,413	16,808	3,739	2,305	17,170
Q^3	87.9	82.4	81.9	82.3	78.1	76.2
Median	68.5	61.2	58.9	57.8	54.9	51.1
Q^1	40.9	35.4	32.3	33.7	29.6	28.0
Q	23.5	23.5	24.7	24.3	24.3	24.1

Psychological percentiles	Semi-skilled	Unemployed and not indicated	Unskilled	Farmers	Total
90–100	429	836	989	1,800	11,191
80– 89	450	804	1,149	1,865	10,436
70– 79	357	770	1,233	1,901	9,537
60– 69	422	861	1,318	2,207	10,188
50– 59	391	803	1,136	2,037	9,010
40– 49	496	1,021	1,590	2,972	11,688
30– 39	394	836	1,317	2,780	9,753
20– 29	365	815	1,262	2,664	9,059
10– 19	438	885	1,502	3,098	10,182
0– 9	366	999	1,505	3,294	9,776
Total	4,108	8,630	13,001	24,618	100,820
Q^3	75.8	73.3	71.0	67.3	76.2
Median	49.9	47.6	45.8	41.5	50.0
Q^1	26.1	23.4	21.9	19.2	25.8
Q	24.8	24.9	24.5	24.0	25.2

Taussig.[1] Because of the large farming population in Wisconsin it seemed advantageous to make a separate grouping for farmer's children. Separate classifications were also made for the professions, for executives, and for business men. A group of miscellaneous and one including those whose parents were unemployed and those who did not indicate the parental occupation were added to present the entire distribution of cases.

Practically all studies similar to this one have found the children of professional parents to rank highest and those of farmers to rank either at or near the bottom of the list. The distribution of mental ability of the high-school seniors in each of the ten parental occupation groups, presented in Table I, lends support to this general conclusion, for the children in the professional group had the highest median mental ability and the children of farmers the lowest median ability. The range of twenty-six points in medians between the farmer group and the professional group is more than one-fourth the range of the entire distribution and the differences in medians between most of the other groups are also great enough to be significant.

TABLE II.—PARENTAL OCCUPATION AND SCHOLASTIC APTITUDE RELIABILITY OF DIFFERENCES BETWEEN MEDIANS. ONE HUNDRED THOUSAND EIGHT HUNDRED TWENTY HIGH-SCHOOL SENIORS

The Figures Represent $D/PE_{diff.}$

	Professional	Executive	Business	Clerical	Miscellaneous	Skilled	Semi-skilled	Unskilled	Farmers
Professional	...	13.46	24.47	18.16	19.32	45.46	32.38	54.49	73.83
Executive	...	.....	4.60	5.20	8.19	19.95	17.32	29.60	40.78
Business	...	.....	.....	2.11	5.93	23.43	16.76	35.82	56.61
Clerical	...	.....	.....	.....	3.49	11.49	11.26	20.79	29.94
Miscellaneous	...	.....	.....	.....	.....	5.20	5.48	12.29	19.17
Skilled	...	.....	.....	.....	.....	.....	3.03	16.34	34.72
Semi-skilled	...	.....	.....	.....	.....	.....	.....	7.36	16.19
Unskilled	...	.....	.....	.....	.....	.....	.....	.....	12.27
Farmers									

To determine whether these differences were reliable, the reliability of the difference between each pair of medians was computed. Table II presents $D/PE_{diff.}$ between the medians of the ten occupational

[1] Taussig, F. W.: *Principles of Economics.* Vol. II, pp. 134–137.

groups. A $D/PE_{diff.}$ of four or higher indicates that the obtained differences between the medians is completely reliable and that a true difference is insured between the samples. It will be seen that all of the differences between medians were found to be sufficiently large to indicate a true difference except between Business and Clerical, between Clerical and Miscellaneous, and between Skilled and Semi-skilled occupations. That is to say, all the differences in the medians, aside from these three exceptions, have a statistical reliability large enough to guarantee that true differences in average mental ability exist between children in the various groups.

However, other facts appear in Table I which seem to be of greater importance in drawing general conclusions from the data. Although the average mental ability of children progresses from one parental occupation group to another, the correlation between mental ability of the student and the rank of the parental occupation, here discovered, for the entire group of students is only +.18. This low correlation is due to the great overlapping from group to group, to the similarly wide range within each group, and to the wide variations of the number of cases in the different occupational groups. There are more than twice as many children of farmers as children of professional men above the median in mental ability. The ranking of the occupational groups listed in order according to the proportion of children each contributed to the total above the median ability is shown in Table III.

TABLE III.—PER CENT EACH GROUP CONTRIBUTES TO THE NUMBER ABOVE THE MEDIAN

Farmers	19.4
Business	19.3
Skilled labor	17.4
Unskilled labor	11.5
Professional	7.9
Unemployed and not indicated	8.0
Executive	5.2
Clerical	4.3
Semi-skilled labor	4.0
Miscellaneous	2.5

From this table it will be seen that the Farmer and Labor groups (skilled, unskilled, and semi-skilled) contributed more than fifty per cent of the cases above the general median in mental ability even though these occupational divisions have the lowest medians. In other words, more than half of the superior children in the high schools of Wisconsin come from the occupational groups which rank lowest.

Table IV presents the separate data for each of the five years. It shows that the rank order of the different occupational groups was constant throughout the five-year period except for the interchange in rank of Executive and Business for 1931, for Business and Clerical for both 1929 and 1933, and for Clerical and Miscellaneous occupations for 1930. These last three shiftings of places account for the lack of statistically significant differences between Business and Clerical, and Clerical and Miscellaneous which was revealed in Table II.

TABLE IV.—PARENTAL OCCUPATION AND MEDIAN SCHOLASTIC APTITUDE PERCENTILE BY YEAR

	1929, median	1930, median	1931, median	1932, median	1933, median	Total, median
Professional	66.5	69.4	69.9	67.8	68.8	68.5
Executive	63.3	63.7	58.0	62.2	59.0	61.3
Business	59.3	60.1	59.1	61.4	55.3	59.0
Clerical	59.9	56.6	54.8	59.2	58.8	57.8
Skilled	52.8	50.2	51.4	50.6	50.3	51.1
Semi-skilled	52.0	50.2	48.7	48.8	50.4	49.9
Unskilled	52.5	45.0	39.0	46.6	46.0	45.8
Farmers	41.3	38.1	43.5	42.1	42.1	41.5
Miscellaneous	57.7	59.6	48.5	55.0	53.8	54.9
Unclassified and unemployed	47.1	48.5	47.4	47.4	47.8	47.6

The hierarchy of specific parental occupations in relation to the mental or scholastic aptitude of the high-school seniors is shown on Table V. This table also shows the percentage of children drawn from each occupational group.

The differences in median mental ability between the children in the upper and those in the lower ranking occupational divisions is enormous. Between the children of professors and those of miners the difference in median ability is fifty points, or half the entire range of percentile points. The medians of the ten or dozen highest groups deviate further than the groups at the bottom of the scale from the median of the entire distribution. That is to say, the high ranking groups exceed the median ability to a greater degree than the low ranking groups fall below it.

The groups at the two extremes of the distribution, however, contribute numbers that are far from equal. The six highest ranking groups with medians all above seventy—Professors, Lawyers, Judges,

TABLE V.—PARENTAL OCCUPATION AND SCHOLASTIC APTITUDE OF ONE HUNDRED THOUSAND EIGHT HUNDRED TWENTY HIGH-SCHOOL SENIORS

	Number of cases	Per cent of total	Q^3	Median	Q^1	Q
Professors	154	.15	93.6	84.6	65.0	14.3
Lawyers and judges	521	.51	92.7	80.8	51.2	20.7
Accountants and auditors	268	.27	90.7	77.1	52.5	19.2
Teachers	669	.66	91.0	74.0	47.5	21.7
Journalists	145	.14	91.8	73.5	50.6	20.6
Bankers and brokers	494	.49	89.8	71.8	45.1	22.3
All professional	6,028	5.97	87.9	68.5	40.9	23.5
Manufacturers	331	.33	88.6	67.9	38.7	24.9
Druggists	251	.25	87.3	67.6	40.3	23.5
Doctors	923	.91	87.1	67.5	38.2	24.4
Executive miscellaneous	1,661	1.65	85.3	67.2	45.4	19.9
Professional miscellaneous	461	.46	86.3	66.9	40.3	22.9
Ministers	721	.71	86.4	66.4	39.8	23.2
Managers	1,139	1.13	84.2	66.3	42.0	21.1
Miscellaneous public officials	163	.17	86.0	63.9	42.3	21.8
Printers	479	.47	86.0	63.3	37.3	24.3
Insurance and real estate agents	1,497	1.48	84.1	62.4	36.2	23.9
Postmasters	203	.20	84.1	61.6	35.5	24.2
Engineers	1,689	1.67	83.7	61.2	32.7	25.4
All executive	4,413	4.37	82.4	61.2	35.4	23.5
Salesmen	3,276	3.25	83.5	61.0	35.2	24.1
Bookkeeper	403	.40	84.5	60.9	40.3	22.0
Mail-carrier	792	.78	81.9	60.7	35.4	23.2
All business	16,808	16.64	81.9	58.9	32.3	24.7
Musicians	91	.09	83.5	58.1	26.7	28.3
All clerical	3,739	3.71	82.3	57.8	33.7	24.3
Miscellaneous business	8,607	8.52	80.9	57.7	30.2	25.3
Miscellaneous clerical	1,165	1.16	83.9	57.0	33.2	25.3
Contractors	1,390	1.38	78.7	56.9	33.1	22.8
Civil service	82	.08	75.0	56.0	30.0	22.4
Bakers	269	.27	77.7	55.6	34.1	21.8
Clerks	1,379	1.37	80.2	55.5	31.7	24.2
Dressmakers	112	.11	77.7	55.5	34.0	21.8
Undertakers-embalmers	142	.14	82.3	55.0	31.0	26.6
All miscellaneous	2,305	2.28	78.1	54.9	29.6	24.3
Railroad engineers	528	.52	79.8	54.4	30.4	24.6
Jewelers	102	.10	81.7	54.3	30.6	25.5
Inspectors	304	.30	80.6	54.3	29.1	25.7
Bricklayers	80	.08	80.0	54.2	22.8	28.5
Retired	717	.71	77.3	54.1	29.0	24.1
Autoworkers	156	.15	80.9	54.0	40.3	20.3
Railroad and street car conductors	323	.32	77.5	53.8	29.2	24.1
Housekeepers-houseworkers	1,357	1.34	80.1	53.7	27.9	26.0
Machinists and mechanics	3,291	3.36	76.3	53.5	28.6	23.8
Miscellaneous	821	.83	76.4	53.3	27.9	24.2
Nurses	135	.13	75.2	53.0	29.0	23.0
Electricians	706	.70	76.5	53.0	28.1	24.2
Miscellaneous skilled workers	3,543	3.65	78.7	52.2	32.6	23.0
Cooks	226	.22	80.1	52.8	27.2	26.4
Shoemakers and cobblers	296	.29	77.3	52.8	28.0	24.6
Laundryworkers	94	.09	73.7	52.3	35.0	19.3
Tailors	389	.39	73.9	51.7	28.8	22.5
Lumbermen	409	.41	77.1	51.7	28.4	24.3

TABLE V.—*Continued*

	Number of cases	Per cent of total	Q^3	Median	Q^1	Q
All skilled	17,170	17.14	76.2	51.1	28.0	24.1
Tinsmiths	123	.12	78.0	51.2	28.6	24.7
Decorators	219	.22	75.0	51.0	30.6	22.2
Plumbers	607	.60	75.5	51.0	26.5	24.5
Barbers	659	.65	73.8	50.3	25.7	24.0
Grocers	589	.58	77.7	50.0	28.9	24.4
Miscellaneous semi-skilled	702	.70	74.9	50.1	26.0	24.4
Grand total	100,820	100.00	76.2	50.0	24.8	25.8
Policemen and sheriffs	472	.47	75.7	50.0	26.1	24.8
All semi-skilled	4,108	4.07	75.8	49.9	26.1	24.8
Foremen	1,613	1.60	74.8	49.5	26.8	24.0
Carpenters	2,937	2.91	74.0	49.0	25.4	24.2
Butchers	482	.48	73.3	48.6	24.0	24.6
Mill and factory workers	1,136	1.17	72.2	47.8	23.6	24.2
Unclassified	8,630	8.55	73.3	47.6	23.4	24.9
Painters	878	.87	70.5	47.6	26.5	21.9
Teamsters	232	.23	69.6	47.0	23.1	23.2
Railroad workers	2,081	2.06	73.1	46.8	23.2	24.9
Laborers	7,943	7.87	70.8	45.8	22.1	24.3
All unskilled	13,001	12.88	70.9	45.8	21.9	24.5
Molders	466	.46	73.4	45.2	23.4	25.0
Miscellaneous farmers	176	.17	68.8	45.1	25.8	21.4
Dairymen	451	.45	72.6	44.5	19.7	26.4
Miscellaneous unskilled	611	.61	65.4	43.5	20.8	22.3
Masons-plasterers	569	.56	72.3	43.4	21.1	25.5
Blacksmiths	510	.51	67.8	43.2	20.7	23.5
Janitors	401	.40	70.4	42.0	17.5	26.4
Millers	162	.16	70.2	42.0	16.3	26.9
All farmers	24,618	24.39	67.3	41.5	19.2	24.0
Farmers	23,991	23.77	67.2	41.5	19.1	24.0
Truck drivers and chauffeurs	411	.41	60.7	40.6	20.7	20.0
Boiler makers	81	.08	66.79	39.2	12.0	27.3
Firemen	385	.38	64.4	39.1	17.9	23.2
Miners	184	.18	62.6	34.2	11.6	25.5

Accountants, Auditors, Teachers, Journalists, Bankers and Brokers—contribute only about two per cent of the total number of children, while the six groups with the lowest medians contributed over twenty-five per cent of the total number of pupils. This supplements the figures presented on Table IV which showed that the groups which had the highest medians did not contribute the greatest number of children above the general median.

From these tabulations the following facts may be concluded:

1. That when seniors in Wisconsin high schools are tabulated according to parental occupation the median mental or scholastic ability differs from one group to another. Most of these differences

are statistically reliable and are highly consistent over a five-year period.

2. That there is great overlapping of scholastic ability within the various parental occupational groups and that the differences within every group are greater than the differences between the groups.

3. That because of the great range within every group and the great variation in the numbers within each group there is a low correlation (+.18) between parental occupation and the mental ability of the pupils.

4. That the groups which rank highest deviate further from the general median of the total group than the lowest groups; that is to say, the lowest groups are not as far below the general median as the highest ranking groups are above it.

5. That because of the similarly wide range of mental ability in all the parental occupational groups and because of the larger numbers of pupils in the lower ranking groups, more than half of the children above the general median are from the Labor and Farmer groups—the two which rank lowest.

EVALUATIVE ATTITUDES AND ACADEMIC SUCCESS

J. W. M. ROTHNEY

Harvard University

The purpose of this paper is to report the results of an experiment designed to determine the value of scores made on a revision of the Study of Values[1] in the prediction of academic achievement of high-school boys. The experiment composed one unit in a series carried out to determine the relationship between scores on various measures of interests and attitudes and academic achievement measures.

The Study of Values which presents a unique approach to the problem of attitude measurement was used to obtain the attitude scores. This scale aims to measure the relative dominance of six basic types of interests or motives in personality postulated by Spranger.[2] The types which he postulates are so elaborate that a complete description of them cannot be given here but a brief characterization of them is presented as follows:

1. Theoretical: The dominant interest of the theoretical man is the discovery of *truth*.
2. Economic: The economic man is characteristically intererted in what is *useful*. He is practical.
3. Aesthetic: The aesthetic man sees his highest value in *form* and *harmony*.
4. Social: The highest value for this type is *love of people*.
5. Political: The political man is interested primarily in *power*.
6. Religious: The highest value for the religious man is *unity*. The experience of unity may be obtained through ordinary religious activities and meditation about religious subjects.

Allport and Vernon, the authors of the original test, point out the way in which an individual's evaluative attitudes may be disclosed in the various activities of everyday life in the statement below.

One may look at a wedding ring, for example, and regard it as an object that is glittering (an aesthetic evaluation), as made of 14-carat gold (theoretical), as a marketable object (economic), as a symbol of love and loyalty (social), as an emblem of rights and duties (political), or as an object with sacramental, mystical significance (religious).

The Study of Values is composed of a series of forty-five questions, the answers to which are interpreted as revealing the relative dominance of the individual's evaluative judgments.

Examination of the language of the original scale indicated that it would present difficulties of interpretation to eleventh-grade boys and in order to test the validity of this supposition, the original scale was worked through with a group of fifteen boys in grades ten, eleven, and twelve. The boys were asked to indicate any words or phrases that they did not understand, to point out any parts of the scale in which the meanings were obscure, and to report any other difficulties encountered in making response to the items. As a result of this primary investigation, changes were made by substituting easier vocabulary and substituting student for adult situations so that only five items retained the exact wording of the original scale.* The nature of the changes may be observed by noting the following items. The first of each pair is taken from the original scale and the second is taken from the revision:

ITEM 1. ORIGINAL ADULT

All the evidence that has been impartially accumulated goes to show that the universe has evolved to its present state in accordance with mechanistic principles, so that there is no need to assume a first cause, cosmic purpose, or God behind it. (*a*) Yes. (*b*) No.

HIGH-SCHOOL REVISION

All the evidence that has been fairly gathered goes to show that the world developed automatically so that there is no need to believe that there was a first cause or God behind it. (*a*) Yes. (*b*) No.

ITEM 2. ORIGINAL ADULT

At an evening discussion with intimate friends of your own sex, are you most interested when you talk about:

(*a*) the "meaning" of life.
(*b*) philosophy and psychology.
(*c*) literature.
(*d*) socialism and social amelioration.

HIGH-SCHOOL REVISION

At an evening discussion with close friends of your own sex, are you most interested when you talk about:

(*a*) the "meaning" of life.
(*b*) what people are and how they behave.
(*c*) literature.
(*d*) socialism and social betterment.

* Professor G. W. Allport assisted in making the changes.

ITEM 3. ORIGINAL ADULT

Do you think a good government should aim chiefly at:
(*a*) more aid for the poor, sick, and old.
(*b*) the development of manufacturing and trade.
(*c*) introducing more ethical principles in its policies and diplomacy.
(*d*) establishing a position of prestige and respect among nations.

HIGH-SCHOOL REVISION

Do you think a good government should aim chiefly at:
(*a*) more help for the poor, sick, and old.
(*b*) the development of manufacturing and trade.
(*c*) introducing more square dealing with other nations.
(*d*) establishing a position of importance and respect among nations.

The complete scale consists of forty-five items requiring one hundred twenty answers, twenty of which refer to each of the six attitudes (Theoretical, Economic, Aesthetic, Social, Political and Religious). The scoring scheme of the original test was used in the revision with a few minor changes. Since the scheme is a rather elaborate one, it will not be presented here, but readers who are interested will find the original scheme in the manual which accompanies the scale[3] and the modification of it in the writer's major study.[4]

Since the original scale and the revision are not similar enough to make exact comparisons between scores obtained on each, it is of interest rather than of significance to note that the high-school students are more political, more religious, less theoretical and less aesthetic than college students. The question is raised as to whether these differences are due to the selection by the colleges or due to changes in attitudes as the result of training received at higher institutions.

The next step in the experimental procedure after the completion of the revision was that of correlating the revision scores obtained by three hundred six eleventh-grade boys in seven high schools with teachers' marks given to the boys in English, Latin, French, Geometry, Algebra and an average of the first four of these over a period of one year. In order to eliminate the influence of scholastic aptitude and chronological age on school achievement the following procedure was utilized. Predictions were made of the students' achievement on the basis of scores on the Terman Group Test of Mental Ability and chronological age. Actual scores were then subtracted from predicted scores to obtain measures of achievement uncorrelated with age and

intelligence test scores. These measures which will be referred to hereafter as *D* (discrepancy) scores are the achievement measures which were correlated with the revised Study of Values scores. The results are reported in Tables I and II.

An examination of Table I reveals that the median of thirty-six coefficients of correlation between the scores on the Values scales and

TABLE I.—COEFFICIENTS OF CORRELATION BETWEEN SCORES ON THE REVISED STUDY OF VALUES AND *D*-SCORES OF ACHIEVEMENT IN SCHOOL SUBJECTS. THREE HUNDRED SIX ELEVENTH-GRADE BOYS IN SEVEN PUBLIC HIGH SCHOOLS OF MASSACHUSETTS

	Theoretical	Economic	Aesthetic	Social	Political	Religious
Average *D*-score[1]	.18	−.06	.08	.03	−.13	−.08
English *D*-score	.10	−.08	.10	.09	−.13	−.05
Latin *D*-score	.12	−.02	−.01	−.01	−.09	.00
French *D*-score	.13	−.02	.06	−.01	−.06	−.08
Geometry *D*-score	.24	−.05	.08	.01	−.13	−.13
Algebra *D*-score	.09	−.01	.09	−.01	.06	−.09

[1] The mean of English, Latin, French, and Geometry marks.

the measures of school achievement is −.01, and that the coefficients range from a maximum of +.24 to a minimum of −.13. The coefficients indicate an efficiency in the prediction of academic achievement of from zero to two per cent over and above the efficiency, for the same purpose, of the best weighted combination of intelligence test scores and chronological ages.

TABLE II.—COEFFICIENTS OF CORRELATION BETWEEN SCORES ON THE REVISED STUDY OF VALUES AND *D*-SCORES OF ACHIEVEMENT IN SCHOOL SUBJECTS. ONE HUNDRED FORTY-NINE BOYS IN THREE PUBLIC HIGH SCHOOLS OF MASSACHUSETTS

	Theoretical	Economic	Aesthetic	Social	Political	Religious
Average *D*-score[1]	−.05	.10	−.09	.06	−.14	.12
English *D*-score	−.01	.03	−.18	.13	−.09	.07
Latin *D*-score	−.09	.03	−.06	.08	−.06	.10
French *D*-score	.00	−.02	.02	.05	−.19	.17
Geometry *D*-score	.13	.01	−.02	−.04	−.09	.02
Algebra *D*-score	.13	.15	−.10	.10	−.07	−.07

[1] The mean of English, Latin, French, and Geometry marks.

A check was made on the findings by repeating the procedure upon a second group of one hundred forty-nine boys. Examination of the results obtained with the second group reported in Table II reveals no marked directional changes, or changes in size of the coefficients. The median of the thirty-six coefficients in the second group is .08 and the range from +.17 to −.19. The differences between the coefficients are small and they appear to be due to the chance fluctuations that may be expected to occur when smaller samplings are tested. All the results indicate quite definitely that there is no substantial relationship between evaluative attitudes, as measured by this method, and school achievement. The greatest efficiency in predicting academic success of this scale is less than two per cent.

In order to determine whether the small correlation coefficients were due to the unreliability of the scale or to the unreliability of the

TABLE III.—RELIABILITY COEFFICIENTS, PROBABLE ERRORS OF THE RELIABILTY COEFFICIENTS, AND STANDARD DEVIATIONS OF THE DISTRIBUTIONS OF SCORES ON THE REVISED STUDY OF VALUES. ONE HUNDRED FORTY-NINE ELEVENTH-GRADE BOYS IN THREE PUBLIC HIGH SCHOOLS IN MASSACHUSETTS

Value	r_1	PE_r	σ
Religious	.60	.035	7.07
Theoretical	.56	.038	5.99
Aesthetic[1]	.51	.041	5.79
Political	.34	.048	6.35
Economic	.30	.051	5.17
Social	.00	.055	4.82
All six	.42	.018	3.80

[1] The first three approximate very closely the reliabilities of the values of the original scale.

criteria of achievement, examinations were made concerning the reliability of each. The reliability of the achievement measures was greater than .80 in all cases. By reference to Table III in which is reported the reliabilities of the Values scales it will be seen that the social Value has zero reliability,* that the political and economic Values approach zero reliability and the religious, aesthetic, and theoretical Values cannot be depended upon for narrow classification. Correction

* Allport and Cantril report: "All evidence goes to show that the measurement of the social value (original scale) is unsatisfactory."

for attenuation of the correlation coefficients between the Values scores and the measures of achievement revealed no significant increase in the size of the coefficients.

The limit of efficiency in the forecasting of achievement imposed by the unreliability of the measures of academic attainment was fifty-eight per cent.* Since the greatest obtained forecasting efficiency of any set of Value scores was less than two per cent, we can be certain that the unreliability of the criteria is not the main factor operating to reduce the forecasting yield. The unreliability of the scale must, therefore, be responsible.

Although the Revised Study of Values seems to have very little practical value in the forecasting of academic achievement, the scores obtained from it reveal some interesting relationships. In Table I, for example, it can be seen that all the theoretical-achievement coefficients are positive, and all the political-achievement coefficients are negative. Similar relationships may be noted in Table II. It appears that theoretical interest is opposed to political interest and that the theoretical person achieves not because he is dominantly theoretical, but because he is theoretical and not political.† Other interesting relationships are to be found in Tables I and II, but it must be pointed out that none of the coefficients are significantly large.

Notwithstanding the deficiency of the Revised Study of Values in forecasting academic achievement, it does appear to have considerable value as a *teaching*, rather than a *measuring*, technique. It may be used for introducing the study of controversial issues in the Social Studies; for introducing the study of personality which might well be brought down to the high-school level with considerable profit to those students who do not go to college; as an exercise in self-rating and self-study; and for introducing students to the study of human motives and conduct. The enthusiastic manner in which students respond to the Revised Study of Values indicates that any high-school teacher would find it to his advantage to give this scale to his students as a classroom exercise.‡ The writer wishes, however,

---

* This figure is obtained by substituting the square root of the reliability coefficient of the achievement criteria in the formula for $E(E = 1 - \sqrt{1 - r^2})$, the per cent of forecasting efficiency.

† There is zero correlation between Revised Study of Values scores, Scores on the Terman Group Test of Mental Ability, and age.

‡ Mimeographed copies of the Revised Study of Values may be had on application to the writer.

to emphasize the fact that the scale cannot be used as a reliable indicator of personality development or scholastic success.

A survey of the literature of educational experimentation reveals that, as in the experiment described above, failure has invariably attended attempts to predict students' marks by means of scores obtained from instruments designed to measure non-intellectual factors of personality. Not one of the many studies reported shows any significant gains over the results obtained by Kelley in his pioneer study in 1914.[5] It appears, therefore, that—reluctant as educators are to give up hackneyed procedures—it will be necessary to discard them until such time as we develop measures of the evaluative attitudes of teachers which determine their distributions of marks. There can be little doubt now that many of the significant factors which determine students' marks lie within the teacher as well as within the pupil. These factors may attenuate the prognostic potency of any instrument so that the scores obtained from it are useless in forecasting marks.

To those who would make *practical* use of any measure of attitudes, interests, and personality so far developed in the prediction of teachers' marks, it will be well to remind them of Allport's[6] pertinent statement:

> It is a mistake at this early stage to confuse research in personality with the ulterior motives of character education, social reform, and industrial exploitation. It is not a service to society to apply psychological theories which are tentative and so clearly destined to revision or total rejection.

BIBLIOGRAPHY

1. Vernon, P. E. and Allport, G. W.: "A Test for Personal Values." *Journal of Abnormal and Social Psychology*, Vol. XXVI, No. 3, 1931.
2. Spranger, E.: *Types of Men.* Translated from the 5th edition of Lebensform. Halle: Niemeyer (American Agent, Stechert), 1928, 402 pp.
3. *Manual of Directions* to accompany the *Study of Values.* Houghton Mifflin Company.
4. Rothney, J. W. M.: *Interests in Relation to School Success.* Doctoral Dissertation, on file Harvard Graduate School of Education Library.
5. Kelley, T. L.: *Educational Guidance; An Experimental Study in the Analysis and Prediction of Ability of High School Pupils.* Teachers College Contributions to Education, No. 71, 1914, 116 pp.
6. Allport, G. W.: Review of P. M. Symond's "Diagnosing Personality and Conduct." *Journal of Social Psychology*, Vol. III, No. 3, 1932, pp. 391–397.

ON THE LOSS OF KNOWLEDGE BY JUNIOR HIGH-SCHOOL PUPILS OVER THE SUMMER VACATION

MARIE SCHREPEL AND H. R. LASLETT

Oregon State College, Corvallis, Ore.

The purpose of this study was that of measuring the amount of forgetting of factual material by junior high-school pupils over a summer vacation lasting three months and two weeks. Unfortunately, there is no way at the present time of measuring the subtler values claimed to result from the work of the schools. Some doubt is thrown on the extent of these subtler values because very few teachers are able to fill out off-hand the information required for even the simplest sociological case histories for any of their pupils. Few teachers know their pupils at all well. Nor can one say that this fault is primarily that of the teachers. In most schools the classes are large and the teaching load heavy. On top of these are the extracurricular activities.

In this study, a preliminary experiment involving fifty-one pupils in arithmetic fundamentals and seventy-four in arithmetic reasoning was carried out in 1933. The Los Angeles Diagnostic Tests in Arithmetic were used. The tests were given twice to the same pupils—once late in April and again during the first week of school in September. As these pupils were in the eighth grade, they were receiving instruction in arithmetic at the time, but no special effort was made, between the April test and the close of school, to stimulate these pupils in the retention of number facts and relations. No mention was made of a second test to be given in the Autumn until the time at which the test was to be administered. No summer-school training was open to any of the pupils involved in either of the experiments in this study. The test records of only those pupils who took both the Spring and the Autumn tests are included in the report of these experiments.

A second experiment, involving the pupils in the eighth and ninth grades of a junior high school, and employing the New Stanford Achievement Test, was carried out in 1934 in much the same way as the earlier experiment had been except that the Spring test was given during the last week in May instead of in April. Form *W* of the Stanford Achievement Test was used in the Spring, and Form *V* in the Autumn. The number of pupils involved was one hundred

twenty-one who completed both forms of this test. Of these, seventy-two were in the eighth grade and forty-nine in the ninth grade. After both forms were scored, the results were tabulated for each of the ten subtests, and the gains or losses of each pupil in all of the subtests were found separately. The means and the standard deviations of the gains or losses on each of these subtests were found. These have been compiled in Table I below:

TABLE I.—A SUMMARY OF THE GAINS OR LOSSES OF KNOWLEDGE OF FACTUAL MATERIAL OVER THE SUMMER VACATION

Test No.	Tests and subtests	Mean change	SD[1] of change
	Los Angeles Diagnostic Tests.		
1	Fundamentals of arithmetic, form 1.....	−6.463	8.73
2	Reasoning in arithmetic, form 2.....	− .135	4.70
	New Stanford Achievement Test, eighth grade: seventy-two cases.		
3	Paragraph meaning.....	− .767	7.56[2]
4	Word meaning.....	+1.733	4.65
5	Dictation.....	+1.164	4.96
6	Language usage.....	+1.379	7.92
7	Literature.....	+2.676	9.96
8	History-civics.....	+ .500	6.88
9	Geography.....	+ .166	6.54
10	Physiology-hygiene.....	+2.885	7.14
11	Arithmetic reasoning.....	+ .052	7.12
12	Arithmetic computation.....	−2.530	8.10
	New Stanford Achievement Test, ninth grade: forty-nine cases.		
13	Paragraph meaning.....	+2.544	5.16
14	Word meaning.....	+2.684	3.84
15	Dictation.....	+ .940	3.66
16	Language usage.....	− .140	6.51
17	Literature.....	+1.385	6.75
18	History-civics.....	+2.859	6.02
19	Geography.....	+ .910	5.76
20	Physiology-hygiene.....	−2.680	8.90
21	Arithmetic reasoning.....	− .245	4.95
22	Arithmetic computation.....	−4.146	8.19

[1] These standard deviations are in no case as much as one-third of their means since the scores on which the means were found extend in both directions from zero.

[2] The original data may be obtained by writing either of the authors at Corvallis, Oregon.

Within each tabulation of the data for each of the subtests, the pupils' scores and their gains or losses were arranged in a rank order according to the mental ages of the pupils as shown by the Otis Self-Administering Test of Mental Ability. This was done to see whether or not the more mature pupils, through greater ease of learning and better comprehension, forgot less than did the less mature pupils over the summer. In all but six of the twenty-two subtests, this was shown to be the case, both in the smaller quantity of loss or greater quantity of gain and in the larger number of pupils who gained or the smaller number who lost in the upper halves and in the uppermost decile groups when compared with the lower corresponding groups. This superiority of the more mature pupils was not clear-cut nor did it hold for all of the pupils in the upper groups, but the data show an obvious tendency in this direction. From a study of the individual pupils involved, it seems probable that the motivation of the less mature pupils was greater than that of the more mature pupils.

The six subtests in which the upper groups lost as much as or more than the lower groups were all in the Stanford Achievement Test series. These tests were: paragraph meaning in the eighth grade; word meaning in the eighth grade; literature in the eighth grade; arithmetic computation in the eighth grade; dictation in the ninth grade; and arithmetic computation in the ninth grade. When the comparison of the uppermost and the lowest decile groups of the eighth and ninth grades together are made, the lower decile groups show greater gains only in the paragraph meaning, the word meaning, and the literature tests. A reasonable explanation of this situation seems to be that an intensive campaign was put on during the school year, 1933–1934, to interest the pupils in reading for pleasure and for general information rather than for the sake of compliance with the program of the school. Most of the effort in this campaign was directed toward the poorer readers who were chiefly in the lower deciles in mental age. Most of the gains for the lowest decile group in the reading fields were made by one boy who said that he had read more books in the summer of 1934 than in all of his life prior to that time. Even in these subtests, however, the scores—in contrast with the gains or losses—of the upper decile group in mental age were distinctly higher than the scores of the lowest decile group. It may be of interest to note that the range of mental age in the eighth grade was from 17-3 (seventeen years and three months) to 9-3, and in the ninth grade from 17-5 to 12-8.

The only loss of knowledge of these junior high-school pupils that is worthy of any consideration was in arithmetic computation. This loss amounted to approximately four and a half months of school work in the eighth grade and to seven months in the ninth grade as shown on the profile sheet of the New Stanford Achievement Test. Whether this is a real loss in terms of revivability after practice in the Autumn and in terms of life usage is a debatable question.

SUMMARY

1. Pupils of the junior high school, as exemplified by the one hundred seventy-two pupils of this study, do not show any serious losses of knowledge of subject-matter over a Summer vacation of fourteen weeks, with the possible exception of knowledge of arithmetic computation.

2. In fourteen out of the twenty-two subtests, these pupils showed actual gains in factual knowledge over the Summer vacation. These were small mean gains, probably due to the maturation of the nervous systems of the pupils and to their free reading.

3. These pupils returned to school in the Autumn with greater reading ability than they had had when they left the school in the Spring.

4. The pupils of greater mental age, as a group, lost less or gained more knowledge over the Summer vacation than did the pupils of lesser mental age. As a group, also, they made decidedly higher scores on the tests than the pupils of lower mental ages. This raises once more the question of the value of speed in education *versus* maturation of the nervous system in pupil grade-placement.

5. The strenuous reviews in which some teachers indulge each Autumn to overcome the loss of knowledge that the pupils are supposed to have sustained over the Summer do not seem to be warranted by the amount of forgetting shown by this experiment. Moreover, these severe reviews are psychologically and pedagogically questionable, if not clearly wrong.

6. Other studies of loss of knowledge over the Summer are appended. Their conclusions about the amounts of forgetting over the Summer are not in complete agreement.

BIBLIOGRAPHY

Brueckner, L. J. and Distad, H. W.: "The Effect of the Summer Vacation on the Reading Ability of First-Grade Children." *Elem. Sch. J.*, Vol. XXIV, 1923–1924, pp. 698–707.

Bruene, Elizabeth: "The Effect of the Summer Vacation on the Achievement of Pupils in the Fourth, Fifth, and Sixth Grades." *J. Educ. Research*, Vol. XVI, 1928, pp. 309–314.

Elder, Harry E.: "The Effect of the Summer Vacation on Silent Reading Ability in the Intermediate Grades." *Elem. Sch. J.*, Vol. XXVII, 1925, pp. 541–546.

Irmina, Sister M.: "The Effect of the Summer Vacation on the Retention of Elementary School Subjects." *The Catholic Univ. of Amer. Educ. Research Bull.*, Vol. I, No.'s 3 and 4, 1928, pp. 99.

Kirby, T. J.: *Practice in the Case of School Children.* New York: T. C. Contrib. to Educ., No. 58, 1913, pp. 96.

Kramer, G. A.: "Do Children Forget During the Summer?" *Baltimore Educ. Bull.*, Vol. VI, 1927, pp. 55–60.

Morgan, L. D.: "How Effective is Specific Training in Preventing Loss Due to the Summer Vacation?" *J. Educ. Psychol.*, Vol. XX, 1929, pp. 466–471.

Morrison, J. C.: "The Effect of the Summer Vacation on Learning." *Ohio State Univ., Educ. Research Bull.*, Oct. 1, 1924, pp. 245–249.

Noonan, M. E.: *Influence of the Summer Vacation on the Abilities of Fifth- and Sixth-Grade Children.* New York: T. C. Contrib. to Educ., No. 204, 1926, pp. 103.

Patterson, V. W.: "The Effect of the Summer Vacation on Children's Mental Ability and on Their Retention of Arithmetic and Reading." *Educ.* Vol. XLVI, 1925, pp. 222–228.

Schroeder, L. J.: "The Effect of the Summer Vacation on Ability in Typewriting." *J. Appl. Psychol.*, Vol. XVIII, 1934, pp. 282–289.

ON THE SO-CALLED CORRELATION COEFFICIENT

MAURICE FRÉCHET

The University of Paris

Professor P. H. Furfey and Mr. J. F. Daly have recently published two interesting papers on the correlation coefficient which I wish to discuss in the present article. I shall first take up the paper entitled, "The Correlation Coefficient as an Index of Relationship" which appeared in the *Journal of Experimental Education*, Vol. III, 1935, pp. 166–169.

The authors rightly draw a sharp distinction between smallness of the so-called correlation coefficient and non-existence (or weakness) of a relationship. Then they add that in order to measure the degree of relationship, two choices are necessary, "(1) the choice of some mathematical function as a standard of perfect relationship, and (2) the choice of some numerical measure of the degree to which the given bivariate distribution departs from this standard."

It is true that this point of view has prevailed in practice and that the use of the correlation coefficient corresponds to the case when the standard function chosen is a linear function. It may also be admitted that this is a rather natural way of solving the problem, provided however that an infinite number of correlation indices corresponding to all possible standard functions be adopted. But there are strong objections to the adoption of a standard function. In most cases when statisticians employ the correlation coefficient they have in view only the existence and the degree of relationship whatever be the shape of this relation.

When we want to know whether there is a relation between the price of wheat and the national income, between the mortality rate and the number of hospitals, between ability in Greek and Latin and ability in mathematics, and so on, it matters very little whether the relation (if it exists) be linear or not. It may sometimes be said that this secondary matter may have no meaning at all, as in the last instance, in which the relation (if it exists) may be linear or not according to the measure of ability employed. Thus a true measure of relationship ought to be independent of the choice of a standard function. As the so-called correlation coefficient is not so, it may be a good coefficient of *linear* relationship, while it is a bad correlation coefficient. This is why I proposed some years ago to call it *coefficient de linéarité*.

Besides, the authors themselves, though they start from the necessity of the choice of a standard function, come to the conclusion above presented, when they say: "The value of the numerical measure of closeness of relationship depends on this choice [of the standard function]. Therefore the *r-eta* definition of closeness of relationship is ambiguous."

From this quotation one may gather that they base their objection to the use of *r* and of *eta* on the same ground. Now it is not easy to see how the objection can apply to *eta*. There is nothing in the motives generally alleged to justify the use of *eta* which presupposes that a standard function has been chosen. It is true that *eta* is still an imperfect measure of correlation since it may be zero for a perfect relationship which is not single-valued; but this is an infinitely broader limitation than the condition of linearity, so that—on the supposition, at least, that we do not know in advance the shape of the relation if it exists—the *eta*-technique is considerably more justified than the *r*-technique.

The authors mention in another sentence the real ground on which the use of *eta* is criticized. They say: "Each bivariate distribution can be made to yield one *r* and several *etas*." This probably means that when several groupings are used, they may give approximately the same value for *r*, but notably different values for *eta*. If this is the objection, it has already been met by the writer in an article in *Barometro Economico*, 1933, pp. 53–58 and 125–130, and in an article which is to appear in the December issue of the *Revue de l'Institut International de Statistique*. The answer is that either, (*a*) the regression is approximately linear and then *r* and *eta* being approximately equal, *eta* will not be much more affected by grouping than *r*, or (*b*) the regression being non-linear, the uniqueness of *r* is not an advantage over *eta*, since in this case *r* has no meaning in regard to closeness of relationship.

Finally, although it is quite right to say that as measures of relationship *r* and *eta* are usually considered the best available, it is still possible to progress on this ground as I have shown in the two articles mentioned in the previous paragraph.

Although I do not agree with the authors on the different points mentioned above, there are in their article many observations with which I fully concur. As to their article "Product-moment Correlation as a Research Technique in Education," published in the *Journal of Educational Psychology*, Vol. XXVI, 1935, pp. 206–211, I can only felicitate them on the opportune nature of their inquiry. It has been

said in different quarters that a renewed discussion of the precautions to be taken in using r would be futile, since everyone is familiar with these precautions. The authors have shown that in such a restricted field of statistics as educational science offers, out of two hundred sixty-eight articles, sixty-three used product-moment correlation and that out of these sixty-three "only seven gave any information about the normality or linearity of their distributions or any data from which such information could be deduced." This shows the sore need of propaganda against the improper use of the so-called coefficient of correlation.

THE SUBJECTIVE ELEMENT IN THE MEASUREMENT OF RELATIONSHIP

PAUL HANLY FURFEY AND JOSEPH F. DALY

Catholic University of America

In an article[1] appearing in this issue of the JOURNAL Professor Fréchet of the University of Paris is kind enough to comment on two papers[2] of ours dealing with the correlation problem. He is quite in agreement with us on the main issue, on the fact, namely, that *r* and *eta* are grievously abused in current practice. This is, indeed, a fact often mentioned by Professor Fréchet in his own numerous statistical papers.

Professor Fréchet, however, raises a question concerning a minor point. In one of our previous papers we had argued that the use of either *r* or *eta* involves the choice of a standard function which is considered, for the problem in question, a criterion of perfect relationship. The numerical value of *r* or *eta* depends on the extent to which the bivariate distribution being studied conforms to this criterion.

Professor Fréchet differs with this statement as far, at least, as *eta* is concerned. He says, "There is nothing in the motives generally alleged to justify the use of *eta* which presupposes that a standard function has been chosen."

It is possible that the disagreement between Professor Fréchet and ourselves is more apparent than real. That depends on the exact meaning which is to be given to the above statement. It may mean (*A*) that there is no single standard function which reappears each time that *eta* is calculated and which differs from time to time by no more than a change in the values of its parameters; or it may mean (*B*) that no standard function at all enters into the calculation of *eta*.

If sense (*A*) is intended, then we would agree quite readily. Such a rigid function (the straight line) is, indeed, implied by the use of *r*, but no single function of this sort is implied by the use of *eta*.

If, however, sense (*B*) is intended, then we would disagree. For every *eta* implies a regression curve which passes through the means of

---

[1] "On the so-called correlation coefficient." *J. Educ. Psychol.*, Vol. XXVII, 1936, pp. 304–306.

[2] "Product-moment correlation as a research technique in education." *J. Educ. Psychol.*, Vol. XXVI, March, 1935, pp. 206–211.

"The Correlation coefficient as an index of relationship." *J. Exper. Educ.*, Vol. III, March, 1935, pp. 166–169.

the successive arrays. If all the points of the scattergram lie on this curve, then the correlation is considered perfect and *eta* equals unity. The further the points lie from this curve, the lower *eta* is. The regression curve, then, serves as a standard and *eta* depends directly for its value on the choice of a regression curve. In a bivariate distribution there are, of course, two standard curves, one for each *eta*.

The weakness of *eta* lies precisely in the necessity of choosing this standard curve. A scattergram can be divided into arrays in a large number of different ways and all these different ways yield, in general, different regression curves and consequently different *etas*. The only way of choosing one of these *etas* as the "true" *eta* is to select the one yielded by the most "reasonable" curve. But this is wholly a matter of subjective judgment. Consequently *eta* is not a truly objective way of measuring closeness of relationship.

Let us illustrate the facts of the last paragraph by an example. Given a scattergram we may divide it up into either very broad or very narrow arrays. In the former case we normally get a smoother regression curve and a smaller value of *eta;* in the latter case, conversely, we normally get a saw-toothed regression curve and a larger value of *eta*. Which of these two *etas* shall we prefer? We shall probably feel that the smoother curve conceals some of the irregularities of the "real" regression curve, say, of the curve which we would obtain if this sampling were repeated an infinite number of times. We shall probably feel that the saw-toothed curve introduces some spurious irregularities which are, in fact, merely due to "chance." We therefore reject both these *etas* and seek one based on some curve lying between these two extreme forms. But how shall we know when we have hit upon the "true" curve and consequently upon the "true" *eta?* In the last analysis the answer to this question is a matter of judgment. The "goodness" of *eta* upon which we finally decide will depend on how good we are at this particular guessing game.

The above criticism applies directly to *eta;* but the same type of reasoning might be used against *r*. In this case, of course, the difficulty does not lie in choosing between different possible values for *r*, but in deciding whether to use *r* at all rather than *eta*. The only basis for this choice is ultimately the decision whether the given bivariate distribution is best represented by a straight line or by some other curve. While there have been techniques proposed for determining whether or not a sample can be considered as having been drawn from a linear

universe, we do not believe that any of them solves this restricted problem in a purely objective fashion.

NOTE BY PROF. FRÉCHET

The preceding article has been kindly communicated to me by the authors before publication.

I may state that the sentence quoted by the authors was intended to be understood as in the sense of (*A*) mentioned in the above article. That is: No standard function has to be chosen *beforehand and independently of the observations*, when justifying *eta*—just contrary to the case of *r*.

The authors' observation on meaning (*B*) are very interesting in explaining the ground for the uncertainty of the value of *eta* according to the grouping. It must be added that no index of correlation may be of significance when to a value of x there corresponds only one observation. So that, on principle, too narrow arrays *cannot* be admitted. This limits the range of the values of *any* given index of correlation which corresponds to different groupings and consequently limits also the weight of the objection to *eta* on this ground.

I may finally refer to my article in the January number of the *Revue de l'Institute Internationale de Statistique* for hints on new indices of correlation (among which is Gini's index of correlation) which I think better than *r* and *eta*.

MAURICE FRÉCHET.

BOOK REVIEWS

K. KOFFKA. *Principles of Gestalt Psychology.* New York: Harcourt Brace and Company, 1935, pp. XI + 720.

The book under review is a great work, probably the greatest systematic psychology that this generation has produced. Koffka's performance is all the more astounding when we take into consideration that he has written it in a foreign language. How many American psychologists could have written a parallel work in German? Surely Gestalt Psychology is fortunate in its exponents. Probably no school of psychology has such capable and scholarly leaders as Gestalt Psychology has in the persons of Koffka and Koehler. These two psychologists, trained in the best traditions of European scholarship, exhibit an erudition that is very pleasing to the reader.

The argument of Koffka's Principles is closely knit and the reader must spend several weeks of hard study to appreciate its flavour to the full. Not a single line can be skipped. But the readers who find philosophical thinking and writing difficult, will have their powers taxed to the utmost to comprehend its subtle thought. Koffka meets his difficulties honestly and squarely. In only one or two places are they glossed over and in these he appeals to the High Priest of the gestalt movement—Wertheimer. If Wertheimer says so it is so.

The gestaltists are a peculiar mixture of philosopher and psychologist, with the emphasis, perhaps, on the side of the philosopher. Their writings appeal more to the philosopher than to the psychologist. Philosophically they are neutral monists who accept perceptual experience as the measure of all that is real and ultimate. This experience occurs in wholes and cannot be analyzed into discrete parts, any more than water, as water, can be analyzed into hydrogen and oxygen. Nor can the parts be built into wholes. They can be brought together in time and space but in no sense can they be combined into one. If these elements are the cause of a further experience, this new experience is entirely new and is never the mathematical summation of the parts.

Koffka starts his treatise by describing psychology as the study of molar behavior (Tolman's term) taking place in a behavioral rather than in a geographical environment. This geographical environment, however, produces the behavioral. But it is in the behavioral environment (and regulated by it) that real behavior takes place. The geo-

graphical environment is the equivalent of reality; the behavioral the equivalent of appearance.

This explanation is more elaborate than that held by the common or garden psychologist. The latter would say that the geographical environment is the one which may be regarded as the same for everybody. We take out of it, as it were, those parts which fit our natures and previous experiences. Human beings are built according to the same general pattern, and it is this fact that enables them to make a unity of the world they live in, that is, to live together as a community. But for the psychologist the interesting thing to observe is the different uses to which different people put the common world. In other words the objective or geographical world can be regarded as constant; the subjective world is variable. However, the variability of the subjective world can only be determined (or judged or measured) by the objective observation of behavior as performed in a geographical world.

Next, Koffka introduces the "field" concept, borrowed, of course, from modern physics. One would naturally expect gestaltists to accept the behavioral environment as the psychological field. But no, they choose instead the physiological field despite the fact that the psychological field is their fundamental explanatory category. The reason for this lies in *isomorphism*. Any actual consciousness, says isomorphism, is identical with the motions of the atoms and molecules of the brain which causes it. Hence psychology must be defined as the study of behavior in its causal connection with the psychophysical field.

The next topic dealt with is more familiar, namely, the division of the behavioral environment into things and framework. Things look as they do because of the field organization to which the proximal stimulus distribution gives rise. All this is comparatively simple to understand, but when Chapter VI is reached, which deals with the constancies of the environmental field, the reader really has to concentrate. Tri-dimensional space and motion which are next dealt with are also fairly difficult to follow. But what shall be said of reflexes, which are understandable in most psychologies! Here Koffka complicates them by introducing the concepts of the Ego and the Executive. The remainder of the book makes less demand upon the reader; it deals with Memory (founded on a trace theory), Learning, Society and Personality.

Scores of experiments find record throughout the book. They are drawn, of course, from the researches of the gestalt school. Many of

these were new to the reviewer, and American psychologists should be grateful to Koffka, as they are taken from sources not readily accessible. But many of these experiments seem to be quite trivial ones; they are disappointing to one who looks for significance in researches. This triviality is understood when we read that they have not been undertaken because of their significance in their own right, but to serve as demonstrations. What an admission! It sounds suspiciously like bolstering up a case. What puzzles the reader is that Koffka as a student reacted vigorously against the trivial experiments of the older psychology—color-mixing and such like things—as they did not seem to deal with the fundamental facts in psychology. Can he really be satisfied with the experiments he reports in this volume?

Justice cannot be done to this important work in a review extending over a page or two. Moreover, the reviewer is not very sympathetic to gestalt psychology although he has tried to keep an open mind and be fair to it. To him gestalt psychology as described in this authoritative volume seems to be the last remaining legacy of the objective idealists. As Gray truly says "There is no difference between an existence which is nothing but thought and one which is nothing but experience. In neither case can the individual get outside of himself to a world of physical existence which has laws of being, independent of his own personal experience. The gestaltist is doomed to a world in which he finds himself the center of things. Not only do the earth and the planets revolve about him, but even the Deity is just one of his experiences."

Those of us who are interested in the development of a scientific and objective psychology are grateful to the gestaltists for helping to clarify the processes of perception and for their insistence upon the fact of the essential unity of the behaving organism (scientific psychologists had carved up the organism rather badly), but against these must be placed the fact that by the introduction of unknown and unknowable *X*'s into their system, they have seriously retarded the progress of scientific psychology. Gestaltists are essentially metaphysicians rather than psychologists. Their writings have closer affinity with those of St. Thomas Aquinas than with the discoveries and viewpoints of modern science. The Gestalt movement bears to psychology the same relationship that pre-raphaelitism bore to painting—it is on the side-branch not on the main stem of development. Koffka's book, therefore, is important, chiefly from the standpoint of the history of psychology. It marks one of the milestones in the history of psy-

chological thought, but it directs the traveller along a side-path, not along the main road. But it will have to be read and taken carefully into account.

In the new edition, which will certainly be called for soon, the numerous typographical errors should be corrected.

PETER SANDIFORD.

University of Toronto.

CHRISTINE P. INGRAM. *Education of the Slow-learning Child.* Yonkers: World Book Co., 1935, pp. XII + 419.

Special education for the subnormal child has developed rapidly during the past quarter of a century. There have been innumerable articles and some books dealing with all phases of such special education, but because of the constant increase in knowledge of the psychology of subnormal children most of the books were soon out of date. Furthermore, the books have seldom been intended for teachers of subnormal children. Therefore there has always been a need for a practical textbook for the use of students and teachers already engaged in this type of special education.

Miss Ingram has most adequately filled this gap in the literature. Every page of her book shows accurate, immediate knowledge of the subnormal child and his problems, as well as an understanding of the classroom teachers' need for knowing about the problems. In Part I there is a description of the psychology of subnormal children and of the objectives to be kept in mind in their education. While the materials here are not complete in a technical sense, they form a more than adequate picture for the classroom teacher. The second section of the book describes in useable detail the organization and management of special classes, and presents definite project outlines as models for the teacher's classroom activities. This description is based on the actual successful practice in the Special Schools of Rochester, N. Y.

There is no doubt that this book will find wide adoption in those schools which offer training for subnormal class teaching. The work will have an equal, or even greater, usefulness for teachers actively engaged in this type of education. One might almost wish there were a law requiring each teacher in subnormal classes, in fact every elementary school teacher, to study this book. C. M. LOUTTIT.

Indiana University.

K. J. HOLZINGER. *Preliminary Report on Spearman-Holzinger Unitary Traits Study.* Chicago: Department of Education, University of Chicago. No. 1, Raw Correlations, Reliabilities, Means and Standard Deviations for Seventy-eight Variables, Using a Sample of one hundred eighteen Cases, 1934, pp. II + 27. No. 2, Intercorrelations from Re-scoring Tests, Correlations with "*g*," Means and Standard Deviations for Boys and Girls, Preliminary Analyses of Speed and Verbal Tests, Using Thorp Data, one hundred eighteen Cases, 1934, pp. 26. No. 3, Raw Correlations, Correlations Corrected for Age, Reliabilities, Means and Standard Deviations for Mooseheart Sample of one hundred Cases, Comparison of Thorp and Mooseheart Reliabilities, Age Correlations and Basic Tetrads for Correlations with "*g*," 1935, pp. 37. No. 4, Factor Patterns and Residual Correlations for Thorp and Mooseheart Data, 1935, pp. 78. No. 5, Introduction to Bi-factor Theory; Solid and Hollow Staircase Patterns for Sets of Data from Mooseheart, 1935, pp. 57.

During the spring of 1933, about one hundred tests were administered to two groups of pupils, one of about seven hundred children in the Mooseheart schools, the other of nearly four hundred children in the Thorp Elementary School at Chicago.

The first of these preliminary reports presents the statistical constants for seventy-eight test variables, using a sample of one hundred eighteen cases from the Thorp population. The mean chronological age of this group was 13.2 years and the standard deviation was 1.5 years. Each of the seventy-eight tests is briefly described in the report.

Reliability coefficients were computed by odds *versus* evens, by alternate forms, or were estimated by average intercorrelations, and are presented for all but two of the tests. These coefficients range from .12 for tests of perseveration to .96 for a test of recognition of form. Many of the coefficients are low.

Correlations of all variables with chronological age are presented and are generally very small, an exception being in the case of the Dynamometer Test. All of the three thousand three intercorrelations among the seventy-eight variables are presented to four places.

The author expresses the wish that those interested in factorial analysis will make use of the material in whole or in part, and by any method desired.

The second report adds some supplementary correlations and analysis based on the same one hundred eighteen cases. Six tests considered measures of Spearman's perceptual *g* were re-scored to eliminate speed and to measure amount done, and correlations of these new variables with each other and with the remaining tests are presented. Correlations of each variable with the *g* of these six tests are also given, and the justification for considering *g* as a factor among the six tests is presented in the tetrads involved. The saturation of the six tests with *g* is shown by *r* values centering around .94, so that the tests appear to be good measures of some common factor.

Analysis of sex differences shows boys significantly superior to girls on four of the six tests of *g*, on measures of mental speed, strength of grip, arithmetic, mechanical ability, free imagination, and general information. Girls excelled boys in speed of writing numbers, skill in interpreting behavior from pictures, quality of composition, musical discrimination, and artistic appreciation.

A technique of analysis using residual correlations is presented, including a definition of residual correlation and a formula for residual *r*.

After the influence of *g* has been removed from the correlations of eight speed tests, two factors called *a* (mental speed) and *m* (motor speed) emerge. A third factor (speed common to *a* and *m*) common to all tests is also apparent, though more than half the observed variance is accounted for by specific factors. Analysis of thirteen verbal tests shows them more largely measures of *v* (verbality) than of *g*, with specific factors again accounting for the larger portion of the observed variance.

It is not entirely clear how the presentation of the forty-five tetrads from the six perceptual tests constitutes a "justification for considering '*g*' as a factor amongst the six tests," when no outside criterion of *g* has been considered. This may partly account for the appearance of an unsigned postcard recently mailed from Chicago and stating that the *g* of Reports 1–4 is the sum of two factors, one much more general than the above, and the other a group factor limited mainly to the visual tests.

The third report concerns chiefly the correlations from a sample of one hundred cases from the main group at Mooseheart, the cases being selected as affording the best measures of the "*o*" (oscillation) factor which had been inadequately tested in the Thorp sample. The mean

chronological age of these children was seventeen years and the standard deviation 1.9 years.

A total of sixty-three tests are reported on, many being the same as those given to the Thorp children. The age factor, which appeared of little importance in the earlier data, appears very significant in these. Intercorrelations among the sixty-three variables are presented both in raw form and corrected for age by the method of partial correlation. Reliability coefficients are likewise given first in terms of raw score and then as corrected for age, although differences between the two values are generally very small.

The grammar and general information tests were given in such a way as to study motivation effect, some of which appeared in the results of the grammar test, but not in the case of the general information test.

For most of the tests the reliabilities appeared considerably higher than in the Thorp data. The author accounts for this by noting the general intellectual superiority of the Mooseheart group.

The fourth report is devoted largely to applications of the residual correlation technique to the data already presented. The tests used included the complete set of sixty-three for Mooseheart and the fifty-seven for the Thorp group, these latter being among the sixty-three Mooseheart variables. From a study of tetrads the author posits *g* as a general factor among one group of variables. Correlations of all remaining variables with *g* are obtained, after which residual correlations, with *g* eliminated, are found. The tetrads from these residual correlations are next obtained for the variables assumed to measure the next factor. These appear to give good evidence of a single factor with negligible group factors, so the correlations of each variable with this second factor are obtained and from these, further residual correlations, as before.

For the Mooseheart data tables are presented to show the residual correlations after each successive factor has been eliminated, the factors being *g*, *v* (verbality), *a* (mental speed), *m* (motor speed), *o* (oscillation), *d* (mechanical ability), *t* (attention). The *o* (oscillation) factor was not studied in the Thorp data.

The analysis results in a factor pattern for the fifty-seven Thorp variables showing specific factors accounting for sixty-three per cent of the total variance, with *g*, *v*, *a*, *m*, *d*, and *t* accounting for fourteen, ten, five, four, two, and two per cent respectively. For the sixty-three Mooseheart variables with age eliminated, the pattern shows specific

factors accounting for fifty-one per cent of the total variance, with *g*, *v*, *a*, *m*, *o*, *d*, and *t* accounting respectively for twenty-one, twelve, five, three, two, two, and two per cent.

After these eliminations the Thorp data showed residual correlations averaging .0017 with a standard deviation of .0655, where the standard error of zero correlation is .09. For the Mooseheart data the mean residual *r* was −.0008 and the standard deviation .0621, where the standard error of zero *r* is .10. For both sets of data, then, the standard deviation of the distribution of final residual *r*'s is smaller than the standard error for $r = 0$. This would appear to indicate that the analysis has been carried farther than is warranted by the reliability of the data; that some of the variance ascribed to named factors should properly be ascribed to chance.

It is apparent from the method of analysis that the error of measurement to which each test is subject is included in the specific factor for that test. In view of the very meagre reliabilities of many of the tests, one is prompted to wonder if the ascription of variance to general factors has not been carried so far that the remaining specifics are themselves largely chance factors. The very low reliabilities of some of the tests lend credence to such a conjecture.

In the fifth report, the author proposes a so-called Bi-Factor Theory of mental factors which appears, he says, not inconsistent with other factor theories. This theory assumes that a set of correlations may be thought of as due to a number of secondary or simple group factors which meet the test of the tetrad criterion. Each of these secondary factors is also thought of as split into two or more proportional parts, one of which is the principal factor and the other the residuum of the secondary factors. If a factor pattern be set up for a group of variables in which the required criterion holds, the principal factor column will have an entry for each variable, and so will the specific factors column. Between these two columns will appear the group or secondary factor columns, in which entries will appear opposite certain variables only. By arranging the order of variables, these secondary factor entries may be made to run diagonally down across the table. For this reason the author refers to the arrangement as a "hollow staircase" pattern.

A method for making the necessary analysis of test data is presented and a resulting hollow staircase pattern for thirty-six variables chosen from the Mooseheart data is shown. This pattern again shows specific factors accounting for half the variance. Nearly thirty per cent of the variance is ascribed to the principal factor, and no secondary

factor appears to account for more than four per cent. In general, the same finding results when the method is applied to all sixty-three tests of the Mooseheart data.

A drastic shortcoming of the method lies in the requirement that the correlations applied must not involve overlap of factors other than the principal factor as determined by the tetrads. Large quantities of very valuable data may thus be impossible of analysis by the method. This requirement also imposes the necessity for computing large numbers of tetrads if the criterion is to be rigorously applied to as many as, say, a score of variables. And after such computation there remains the difficult if not often impossible task of grouping the variables so as to have each test assigned to only one secondary factor.

Another disturbing feature of the analysis is the finding that the tests earlier shown to be excellent measures of *g* carry relatively small weights for the *g* factor in the hollow staircase pattern. The unsigned postcard from Chicago revises the names of the factors, pointing out that the principal factor is really *g* and that these relatively small weights are for a group factor in the visual tests.

Of course it must be remembered that the "hollow staircase" is not entirely hollow for any but artificial data, since the factor weights which could be placed in the blank spaces may mostly be negligible, but not all equally so, and certainly not all equal to zero. In this connection it is disturbing to observe that the entries in each line of the hollow staircase factor pattern sum to 1.000 in the Report. Have all the "negligible" factor weights been moved into the "specifics" column? If so, we should like to be assured not only of the negligibility of each weight, but also of the negligibility of their sum.

One cannot fail to be impressed by the amount of very careful work which must have gone into the preparation of the five reports. The computations have evidently been performed with many checks for accuracy, and all of the statistical constants are presented to enough places so that their use in further analysis may proceed with confidence.

P. J. Rulon.

Harvard University.

Ben Karpman. *The Individual Criminal.* Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934, p. 317.

Dr. Karpman's latest publication, *The Individual Criminal,* is a study in the psychogenetics of crime. Five detailed case studies serve as a foundation for this study. In a descriptive fashion these

cases have already been presented in very lengthy form in a previous publication called *Psychopathology of Crime.* In the present study, which is intelligible without a reading of the former, the emphasis is on interpretation, differential diagnosis and a consideration of crime mechanics. The case studies proper then are used as a point of reference for consideration of the psychological and social meanings of crime. In his consideration of the specific diagnoses of his case studies the author not infrequently gets involved in rather lengthy discussions concerning the nature of psychoses, neuroses, psychopathy, egopathy, thinking processes, emotional maturity and other such problems.

Many American psychologists, particularly educational psychologists, will probably disagree with the author's treatment of some topics. For example, clinical psychologists might well disagree with his hard and fast differentiation of neuroses and psychoses. As a matter of fact even the author is not over-consistent in this differentiation once he makes it. In his discussion of the first case presented he makes such arbitrary statements concerning neuroses and psychoses as follows: They "represent two entirely different types of reaction, expressions of widely different types of personalities . . . The one can never pass into the other. An extreme and even very fulminating neurosis does not pass into a psychosis, but still remains a neurosis; while even a benign form of psychosis begins not as a neurosis, but as a psychosis . . . The neurotic may either laugh or cry, perhaps out of proportion to the situation, but nevertheless qualitatively in a way demanded by the situation; the psychotic, however, unable to synthesize his emotions, will, at one and the same time, laugh with one side of the face and cry with the other, showing marked disharmony, or laugh without any adequate stimulus and in situations in which the opposite emotional reaction would be expected."

Many criminologists and social philosophers will doubtless agree with the author that criminality is one of the greatest social problems today. They probably will also agree with him that an understanding of its history can yield insights about the nature of human culture, that the devising of means to deal intelligently with criminals would be a good way of helping build a happier and better state of society. All of this, however, they would do on a basis that probably would not be in agreement with the assumption made by the author, namely, that "we are born selfish, hateful, spiteful, mean; and it is culture that makes us devoted, loving, kind and sympathetic." These infrequent

discrepancies, however, do not in any way diminish the usefulness of the book. The five case studies are comprehensively and well presented and a short chapter on the psychotherapy of crime is also included. No serious student of criminology can afford to neglect the present volume of Dr. Karpman's.

H. MELTZER.
Psychological Service Center, St. Louis.

JOHN J. B. MORGAN. *Keeping a Sound Mind.* New York: The Macmillan Company, 1935, p. 440.

Keeping a Sound Mind Dr. Morgan has written as a basal text for courses in mental hygiene and also for use in freshman hygiene courses. Because of this its organization and style occasionally are somewhat text-bookish, but the subjects considered therein are treated very directly, dynamically, relevantly and effectively. Of the many popular books on mental health recently written this one is the most readable and, judging from the experiences of the reviewer, will probably serve many people who want to help themselves through reading. No attempt at scholarship or comprehensiveness is made here. The book avowedly aims to give people information and to teach them how to acquire attitudes that will help them maintain healthy minds. It is not a profound book but it is a useful book. Some idea of the nature of it can be obtained by the topics as well as the manner of treating them. Topics considered are: Mental conflicts, the mastery of fear, what to fight for, how to fight, emotional maturity, counteracting defects, overcoming emotional depressions, how to get things done, getting along with people, and self-confidence.

H. MELTZER.
Psychological Service Center, St. Louis.